MEASUREMENT OF THE TOP QUARK MASS IN THE DILEPTONIC DECAY CHANNEL AT CMS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Nathan Mirman

January 2017

MEASUREMENT OF THE TOP QUARK MASS IN THE DILEPTONIC DECAY CHANNEL AT CMS

Nathan Mirman, Ph.D.

Cornell University 2017

This dissertation presents a measurement of the top quark mass (M_t) in the dileptonic $t\bar{t}$ decay channel using data from proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$ recorded by the CMS experiment at the LHC, corresponding to an integrated luminosity of $19.7 \pm 0.5\ \text{fb}^{-1}$. The analysis is based on three observables whose distributions are sensitive to the value of M_t. The $M_{b\ell}$ invariant mass and M_{T2}^{bb} 'stransverse mass' observables are employed in a simultaneous fit to determine the value of M_t and an overall jet energy scale factor (JSF). In a complementary approach, the M_{T2}-assisted on-shell reconstruction technique is used to construct an $M_{b\ell\nu}$ invariant mass observable that is combined with M_{T2}^{bb} to measure M_t. The shapes of the observables, along with their evolutions in M_t and JSF, are modeled by a non-parametric Gaussian process regression technique. The sensitivity of the observables to the value of M_t is investigated using a Fisher information density function. The top quark mass is measured to be $172.22 \pm 0.18\ (\text{stat})\ ^{+0.89}_{-0.93}\ (\text{syst})$ GeV.

This dissertation also presents a missing transverse momentum ($\vec{\not{E}}_T$) significance variable, which is used to estimate the compatibility of the reconstructed $\vec{\not{E}}_T$ with a zero nominal value. This variable may be used to discriminate between events containing real $\vec{\not{E}}_T$ due to undetected particles and spurious $\vec{\not{E}}_T$ due to object misreconstruction, finite detector resolution, or detector noise. The $\vec{\not{E}}_T$ significance variable is tuned using data-driven techniques, and its performance is evaluated using the CMS Run 1 and Run 2 datasets.

BIOGRAPHICAL SKETCH

Nathan Mirman was born in Chicago, Illinois, and later moved with his family to Minneapolis, Minnesota to complete his primary and secondary education. In the spirit of a good Eastern-European child, Nathan spent much of his younger years perfecting the art of piano and chess, while daydreaming about the mysteries of the universe. As a high school student, Nathan was given the opportunity to take physics and math classes at the University of Minnesota through the PSEO program. He found the experience to be very rewarding, and decided to pursue this course of study as an undergraduate.

Nathan earned a Bachelor of Science *summa cum laude* with high distinction at the University of Minnesota in 2010. There, he joined the Minnesota CMS group, where he worked with Professor Jeremiah Mans on electromagnetic triggering in the CMS forward calorimeter endcaps, and contributed to a study of the Z boson rapidity spectrum using first data at CMS. At the University of Minnesota, Nathan discovered his passion for particle physics, and was excited to get involved in the analysis of first LHC collisions as a future graduate student.

Nathan began his Ph.D. work at Cornell in 2010, just months after the start of the first stable run at the LHC. There, he had the privilege to continue his work on the CMS experiment under the direction of Professor Jim Alexander. He was involved in two measurements of the top quark mass, the development of a missing transverse energy significance variable, and pixel detector upgrade studies. In the summer of 2011, and during the years 2014-2015, Nathan was stationed at CERN near Geneva, Switzerland, where he supplemented his research program at CMS with the rigorous study of wine and fondue. Nathan's time at Cornell was defining, both personally and professionally. He was extremely fortunate to be surrounded by a group of incredible people, many of whom will remain lifelong friends.

For my family

ACKNOWLEDGEMENTS

The adventure culminating in this dissertation would not be possible without the insight, inspiration, and support of many incredible people I have encountered along the way.

First and foremost, I would like to thank my advisor, Jim Alexander. It has truly been a pleasure and a privilege to have Jim as a mentor over the last six years. Working with Jim has instilled in me a minimalist approach to research – build the tools you need from the ground up; employ clear, simple methods that highlight the underlying physics; communicate clearly, with ample visual aids; pay attention to detail while keeping sight of the big picture; never use Beamer; always write a toy Monte Carlo. Jim has provided much-needed guidance with utmost subtlety, through gentle nudges, suggestions, and encouragements. And Jim's patience has afforded me the ability to take risks and make a plethora of mistakes, in response to which he has always reacted with reassurance and support. For all of this and a lot more, I have been extremely fortunate to have Jim as my advisor.

It has been a great privilege to be part of the Cornell CMS group over the past several years. Thank you to Peter Wittich for all of his advise, both inside and outside his role as a member of my committee. His friendly, dynamic, candid style, and his sense of humor, have been greatly appreciated over the years. Thank you also to Ritchie Patterson, Anders Ryd, and Julia Thom for fostering a great research environment. In the theory group, I am grateful to Maxim Perelstein for serving on my committee, and for his collider physics seminar, which I found to be very illuminating. Thank you also to Csaba Csáki for his phenomenal lectures on QFT, and Yuval Grossman for his casual and insightful seminars on flavor physics.

One of the best parts of being a graduate student at Cornell is the fantastic group of comrades in arms with whom to weather the storm. A big thanks to

my friends and members of the Cornell CMS community: Avishek Chatterjee, Jorge Chaves, Yangyang Cheng, Jen Chu, Jared Claypoole, Adam Dishaw, Susan Dittmer, Nic Eggert, Ben Kreis, Kevin McDermott, Ben Nachman, Stephen Poprocki, Dan Quach, Aurelijus Rinkevicius, Emmanuele Salvati, André Sterenberg Frankenthal, Zhengcheng Tao, Shao Min Tan, Josh Thompson, Jordan Tucker, Yimin Wang, and Margaret Zientek. It has been a blast working with everyone, traveling, attending boring meetings, attending exciting meetings, driving to and from Point 5, fixing, running, and upgrading CMS, and much, much more. Theorists, it has been a pleasure: Jack Collins, Jeff Dror, Salvator Lombardo, Nic Rey-Le Lorier, and Mike Saelim. And finally, thank you to my good friends from the broader physics world: Daniel Citron, Bob De Alba, Sergei Dyda, Jen Grab, François Hébert, Ethan Kassner, Oleg Kogan, Brian Leahy, Dave MacNeill, Katie McGill, and Veronica Pillar. Friends and physicists – I consider myself very fortunate to have known you all.

At CERN, I have had the chance to work with many great physicists from around the world. Thank you to the top quark experts – Andrea Giammanco, Andreas Meyer, Martijn Mulders, Benjamin Stieger, Pedro Silva, and Rebeca Gonzalez Suarez – whose insight has been much appreciated during the course of our top quark mass analysis. Thank you also to the MET experts – Maria D'Alfonso, Zeynep Demiragli, Phil Harris, Matthieu Marionneau, and Robert Schoefbeck – who have been amazing collaborators.

Finally, thank you to my family. My parents, Julia and Jacob Mirman, and my sister, Hannah Mirman, have been incredibly supportive as I navigated the roller coaster of graduate school and research. And my grandparents, Sophia and Marat Katz, Julia and Isaac Mirman, have lent their support and incisive wisdom, which at times has had a direct impact on my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Nature uses only the longest threads to weave her patterns, so each small piece of her fabric reveals the organization of the entire tapestry.

– Richard Feynman

What are the elementary constituents of matter, and how do they interact? Much of what we know about the subatomic world is encoded in the Standard Model (SM) of particle physics. The matter content of the SM includes three generations of quarks and three generations of leptons. The first generation – containing the up and down quarks, the electron, and the electron neutrino – accounts for most of the stable baryonic matter that makes up the tangible world. The strong force confines these quarks into protons and neutrons, which bind together to create atomic nuclei. Atoms are formed by the electromagnetic interaction of nuclei with electrons, laying the foundation for chemistry. The weak interaction facilitates processes such as nuclear beta decay, a crucial step in the chain of reactions that fuel the burning of stars.

The second and third generations of the SM include quarks and leptons that are more massive, and therefore unstable. They are typically observed in controlled experiments, often with the help of high energy particle accelerators to create them and specialized detectors to reconstruct their decay products. The top quark is the most massive member of the SM, with a mass of roughly 200 times that of the proton, and a short lifetime of 5×10^{-25} s. It has been produced at only two locations – the Tevatron at Fermilab, where it was discovered in 1995, and the

Large Hadron Collider (LHC) at CERN, where further studies of its properties are now underway. Despite its ephemeral nature, the top quark plays a crucial role in tests of the SM and constraints on possible physics beyond the SM (BSM).

The SM is based on the framework of quantum field theory (QFT), where particles are viewed as excitations of quantum-mechanical fields. In this description, the properties of the elementary particles – their masses and lifetimes, as well as their modes of interaction, production, and decay – are determined by the underlying fields, and are subject to constraints imposed on the fields. Central to the SM is the concept of *local gauge invariance*, which requires the theory to be invariant under local transformations of a symmetry group. Imposing this symmetry condition fixes all the possible interactions in the theory – the strong and electroweak forces, as well as the g, $W^{\pm}$, Z, and A bosons that mediate these forces, are entirely specified by the SM symmetry group. Just as symmetry is responsible for dynamics in the SM, *symmetry breaking* results in the generation of mass. The SM weak bosons are massive because they correspond to a broken symmetry, while the gluons and photon correspond to unbroken symmetries and are massless. The Higgs boson associated with symmetry breaking in the SM was postulated in the 1960s, and was first observed at the LHC in 2012.

In this description, the rich phenomenology of particles and interactions in the SM stems from a single set of underlying principles. These principles define relationships between SM observables, and can be probed by precise measurements of the observables. At the same time, the properties of all SM objects are intertwined by quantum-mechanical effects, such as loops and interference. Through such mechanisms, an object as ephemeral as the top quark can have a tangible effect on the masses and decay widths of other particles. Most notable is the con-

nection between the top quark and the Higgs boson, which is often produced at the LHC through top quark loops or in association with top quarks, and can decay through top quark loops to lighter objects. Similarly, the W boson mass carries a dependence on the top quark mass through quantum loop corrections. For this reason, the top quark, W boson, and Higgs boson masses are important parameters in global fits that test the self-consistency of the SM. Quantum loop effects are also responsible for the SM hierarchy problem, where quadratic loop corrections to the Higgs boson mass due to the top quark and other fermions can be of order the Planck scale. The fact that the Higgs mass is 16 orders of magnitude below the Planck scale presents a significant dilemma, and it is a strong motivator for a class of BSM physics models.

This dissertation presents a measurement of the top quark mass in the dileptonic $t\bar{t}$ decay channel, based on a dataset recorded by the Compact Muon Solenoid (CMS) detector positioned on the LHC. Chapter 2 gives an overview of the SM, and outlines topics that are relevant to the top quark. A description of the experimental apparatus, which includes the LHC and the CMS detector, is given in Chapter 3. This chapter also includes a description of the Pixel tracker phase-1 upgrade, which is scheduled to be inserted into CMS in the coming months. Chapter 4 gives an overview of object reconstruction at CMS. Chapter 5 describes the missing transverse energy significance variable, developed using the 8 TeV dataset and commissioned for Run 2 data. The top quark mass analysis is described in Chapters 6-8. Chapter 6 outlines a Gaussian Process (GP) regression technique for nonparametric distribution shape estimation. The core of the top quark mass measurement is described in Chapter 7. Chapter 8 features a sensitivity study, based on a Fisher information density function, of the observables used in the top quark mass analysis. Conclusions and future prospects are discussed in Chapter 9.

CHAPTER 2

THEORETICAL BACKGROUND

Every cultural person must know the Feynman path integral.

– Yuri Orlov

The theory of the top quark is rooted in the Standard Model (SM), which describes the known elementary particles and their interactions by the strong, weak, and electromagnetic forces [31–33]. Since its development in the 1960s and 1970s, the SM has stood up to a variety of experimental tests, and has generated a series of fruitful predictions. In this chapter, we outline the construction of the SM and discuss some of its phenomenological features. We touch on aspects of the SM relevant to the top quark, and motivations of the precise determination of the top quark mass.

2.1 The Standard Model

The SM is based on the local gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$, corresponding to the strong and electroweak interactions. This symmetry group acts on the SM matter fields specified in Table 2.1. The spin-$\frac{1}{2}$ matter fields consists of three generations of leptons,

$$\begin{pmatrix} e \\ \nu_e \end{pmatrix}, \quad \begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}, \quad \begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix},$$

and three generations of quarks,

$$\begin{pmatrix} u \\ d \end{pmatrix}, \quad \begin{pmatrix} c \\ s \end{pmatrix}, \quad \begin{pmatrix} t \\ b \end{pmatrix},$$

written here as $SU(2)$ doublets. Leptons interact by the electroweak force, and quarks interact by both the electroweak and strong forces. The Higgs field, ϕ, is also charged under the electroweak interaction, and attains a nonzero vacuum expectation value (vev) due to the shape of its potential. By this mechanism, the Higgs is responsible for the spontaneous symmetry breaking (SSB) of the electroweak symmetry down to electromagnetism: $SU(2)_L \times U(1)_Y \to U(1)_{\mathrm{EM}}$. The SM matter and boson fields become massive as a result of SSB.

Interactions in the SM are mediated by twelve spin-1 bosons. The spin-1 boson fields are not included explicitly into the theory, but are derived by imposing the principle of local gauge invariance under $SU(3)_C \times SU(2)_L \times U(1)_Y$. The bosons correspond to generators of the SM gauge group – eight gluons generate $SU(3)_C$, and four electroweak bosons generate $SU(2)_L \times U(1)_Y$. After SSB, the electroweak sector is split into the weak and electromagnetic interactions. The former corresponds to the spontaneously broken symmetry, and is mediated by the massive W^+, W^-, and Z^0 bosons. The latter corresponds to the unbroken symmetry $U(1)_{\mathrm{EM}}$, and is mediated by the massless photon, A. Eight gluons, g, correspond to the unbroken $SU(3)_C$ symmetry and also remain massless after SSB.

The SM is entirely specified by the gauge structure, pattern of SSB, and matter content outlined above. In Section 2.1.1, we construct the SM Lagrangian density, or 'Lagrangian', $\mathcal{L}_{\mathrm{SM}}$, by including all allowable terms based on these ingredients. The final Lagrangian has 19 free parameters, corresponding to particle masses, couplings, angles, CP-violating phases, and the value of the Higgs vev. With the

Table 2.1: The matter and Higgs boson field content of the SM. The quarks (Q, u_R, d_R) and leptons (L, e_R) have three generations in the SM. The Higgs field (ϕ) is responsible for SSB in the electroweak sector.

Field	$SU(3)_C$	$SU(2)_L$	$U(1)_Y$
$Q = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$	**3**	**2**	$\frac{1}{3}$
u_R	**3**	**1**	$\frac{4}{3}$
d_R	**3**	**1**	$-\frac{2}{3}$
$L = \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}$	**1**	**2**	-1
e_R	**1**	**1**	-2
$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$	**1**	**2**	1

discovery of the Higgs boson in 2012, existing data can now overconstrain the SM. Precision measurements of SM observables can test the self-consistency of the theory, and probe for hints of BSM physics.

2.1.1 The SM Lagrangian

The Lagrangian $\mathcal{L}_{\text{SM}}$ is constructed by including all terms that satisfy a set of general principles. All terms must be renormalizable to ensure that the theory is specified by a finite number of parameters. The Lagrangian must be invariant under the Poincaré group, which includes translations in space and time, as well as

Lorentz rotations and boosts. This is the full symmetry group of special relativity, and by Neother's theorem, invariance under this group ensures the conservation of energy, momentum, and angular momentum. The Lagrangian must contain the set of matter fields outlined in Table 2.1, with their specified representations under the SM gauge group. Finally, $\mathcal{L}_{\text{SM}}$ must satisfy the requirement of local gauge invariance under $SU(3)_C \times SU(2)_L \times U(1)_Y$, which is spontaneously broken down to $SU(3)_C \times U(1)_{\text{EM}}$ by the SM scalar Higgs field. These conditions specify the dynamics of the SM – characterized by the strong, weak, and electromagnetic forces – and allow the boson and fermions to acquire nonzero masses.

The terms of $\mathcal{L}_{\text{SM}}$ can be grouped into the kinetic, Yukawa, and Higgs sectors:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{kin}} + \mathcal{L}_{\text{Yuk}} + \mathcal{L}_{\phi}. \tag{2.1}$$

The kinetic terms describe the propagation and interactions of the fermions and gauge bosons. The Yukawa terms define the coupling of the SM fermions and gauge bosons to the scalar Higgs field, and the generation of mass terms as a result of SSB. The potential that is responsible for SSB is contained in the Higgs sector. The components of $\mathcal{L}_{\text{SM}}$ are described in more detail below.

Kinetic sector

To construct the kinetic terms of $\mathcal{L}_{\text{SM}}$, we begin with the Dirac Lagrangian, given by:

$$\mathcal{L}_{\text{Dirac}} = \overline{\psi}(i\partial\!\!\!/ - m)\psi, \tag{2.2}$$

where ψ is a Dirac spinor, m is a mass parameter, and $\partial\!\!\!/ = \gamma^{\mu}\partial_{\mu}$. These terms describe a free Dirac fermion, which propagates through space and time but does not interact. Interactions emerge by requiring that $\mathcal{L}_{\text{Dirac}}$ be invariant against

local gauge transformations under a specified gauge group. To demonstrate how interactions arise from local gauge invariance, we consider the local transformation of $\psi(x)$ under the $U(1)$ group, given by:

$$\psi(x) \to e^{i\alpha(x)}\psi(x), \tag{2.3}$$

where $\alpha(x)$ is a function of the spacetime coordinate x. It is evident that the kinetic (derivative) term of $\mathcal{L}_{\text{Dirac}}$ is not invariant under such a transformation. However, it can be made invariant by introducing a vector field A_μ with the transformation property:

$$\mathrm{A}_\mu \to \mathrm{A}_\mu - \frac{1}{e}\partial_\mu\alpha(x), \tag{2.4}$$

where e is a coupling constant. This vector field is attached to the derivative in $\mathcal{L}_{\text{Dirac}}$, resulting in the covariant derivative given by:

$$D_\mu = \partial_\mu + ie\mathrm{A}_\mu. \tag{2.5}$$

Promoting the derivative in $\mathcal{L}_{\text{Dirac}}$ to the covariant derivative and adding a term corresponding to the free propagation of the vector field A_μ gives rise to the QED Lagrangian:

$$\mathcal{L}_{\text{QED}} = \overline{\psi}(i\not{D} - m)\psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{2.6}$$

where $F_{\mu\nu} = \partial_\mu\mathrm{A}_\nu - \partial_\nu\mathrm{A}_\mu$ is a field strength tensor. It is evident that the local gauge transformations in Eqs. 2.3 and 2.4 leave $\mathcal{L}_{\text{QED}}$ unchanged. In addition, the Lagrangian now has terms of the form $e\overline{\psi}\mathrm{A}_\mu\psi$, generating interactions between the fermion fields that are mediated by the gauge boson A.

Before continuing to construct the kinetic terms of $\mathcal{L}_{\text{SM}}$, it is worth pausing here to appreciate the significance of the formulation above. In QED, the fermion field ψ corresponds to an electron with mass m and electric charge e. Starting with $\mathcal{L}_{\text{Dirac}}$, which describes the free electron field, we have derived the photon

field A_μ from first principles by requiring the condition of local gauge invariance. The Euler-Lagrange equation for A_μ yields the inhomogeneous Maxwell equations,

$$\partial_\mu F^{\mu\nu} = e\overline{\psi}\gamma^\nu\psi = ej^\nu, \tag{2.7}$$

with the current density j^ν corresponding to a conserved vector current. The Lagrangian $\mathcal{L}_{\text{QED}}$ describes the properties and interactions of electrons and photons. This remarkably simple two-parameter Lagrangian has been experimentally tested to a precision of up to ten parts *per billion* in measurements of the fine structure constant α, and can account for observed phenomena down to lengthscales of 10^{-15} m. The theory of QED is among the most successful of twentieth century physics.

The kinetic terms of $\mathcal{L}_{\text{SM}}$ are constructed by applying the principles outlined above to the full SM matter content specified in Table 2.1 under the gauge group $SU(3)_C \times SU(2)_L \times U(1)_Y$. The SM covariant derivative is given by:

$$D_\mu = \partial_\mu + ig_s\lambda_a G_{a\mu} + ig\tau_b W_{b\mu} + ig'yB_\mu, \tag{2.8}$$

where $G_{a\mu}$, $W_{b\mu}$, and B_μ are the $SU(3)$, $SU(2)$, and $U(1)$ gauge bosons, respectively; λ_a, τ_b, and y are the generators of these groups; and g_s, g, and g' are the corresponding coupling constants. The indices a and b run over eight gluon and three electroweak gauge fields, respectively. The terms of $\mathcal{L}_{\text{kin}}$ are given by:

$$\begin{aligned}\mathcal{L}_{\text{kin}} =& i\overline{Q}_i \not{D} Q_i + i\overline{u}_i \not{D} u_i + i\overline{d}_i \not{D} d_i \\ &+ i\overline{L}_i \not{D} L_i + i\overline{e}_i \not{D} e_i \\ &- \frac{1}{4} G_{a\mu\nu} G_a^{\mu\nu} - \frac{1}{4} W_{b\mu\nu} W_b^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu},\end{aligned} \tag{2.9}$$

where $G_{a\mu\nu}$, $W_{b\mu\nu}$, and $B_{\mu\nu}$ are the field strength tensors corresponding to the $SU(3)$, $SU(2)$, and $U(1)$ gauge fields, respectively.

From Eq. 2.9, it is evident that mass terms of the form

$$\Delta\mathcal{L} = -m\overline{\psi}\psi = -m(\overline{\psi}_L\psi_R + \overline{\psi}_R\psi_L) \tag{2.10}$$

are not included. Here we have split the Dirac fermion ψ into its right- and left-handed components. Because the electroweak interaction is chiral, these components do not have the same representation under $SU(2)$ and have different charges under $U(1)$ – the Dirac mass term in Eq. 2.10 is in fact forbidden. At the same time, fermions in the SM are known to have nonzero masses. This dilemma motivates additional terms in $\mathcal{L}_{\text{SM}}$ containing a scalar Higgs field. These terms are included in the Higgs and Yukawa sectors, outlined below.

Higgs sector

The Higgs sector is given by:

$$\mathcal{L}_\phi = (D^\mu\phi)^\dagger(D_\mu\phi) - V(\phi), \tag{2.11}$$

where ϕ is a complex scalar field, and D_μ is the gauge covariant derivative defined in Eq. 2.8. The Higgs potential contains the terms:

$$V(\phi) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2, \tag{2.12}$$

where μ and λ are constants. We have separated the kinetic and potential terms of $\mathcal{L}_\phi$ to highlight the fact that the ground state of $\mathcal{L}_\phi$ does not correspond to $|\phi| = 0$. In fact, the potential $V(\phi)$ is minimized at:

$$v = \left(\frac{\mu^2}{\lambda}\right)^{1/2}, \tag{2.13}$$

where $\mu^2 > 0$, $\lambda > 0$, and v is the Higgs vev. The shape of this potential is illustrated in Fig. 2.1. Given an appropriate choice of μ^2 and λ, the ground state of $\mathcal{L}_\phi$

corresponds to a nonzero Higgs vev, thus spontaneously breaking the electroweak symmetry of the SM. As a result of SSB, the kinetic term of $\mathcal{L}_\phi$ picks up mass terms for the electroweak bosons – W^+, W^-, and Z^0 – corresponding to the broken $SU(2)$ symmetry. These mass terms are discussed in more detail in Section 2.1.2. The fermions acquire mass terms from the Yukawa sector, outlined below.

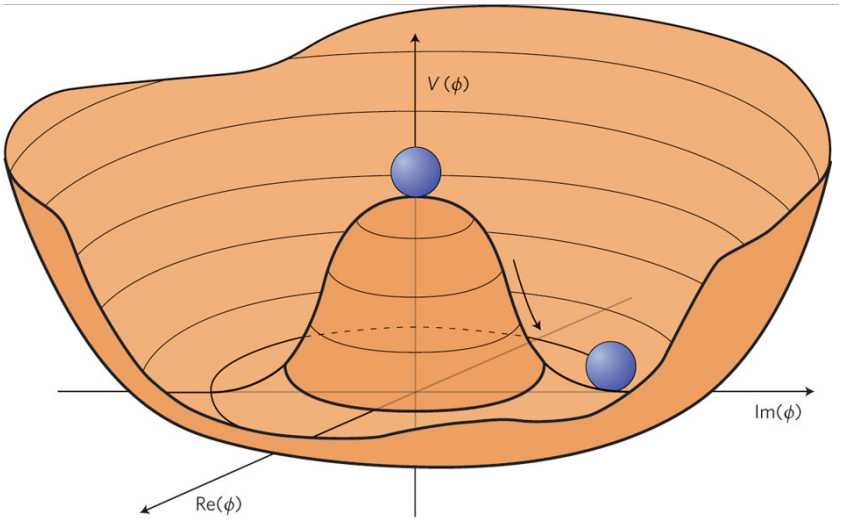

Figure 2.1: When $\mu^2 > 0$, $\lambda > 0$ in Eq. 2.12, the ground state of the Higgs potential does not coincide with the origin. Instead, it lies on a radius of $v = (\mu^2/\lambda)^{1/2}$ around the origin, thus generating a nonzero Higgs vev [2].

Yukawa sector

The Yukawa sector includes interactions between fermions and the scalar Higgs field. It is given by:

$$\mathcal{L}_{\text{Yuk}} = Y^u_{ij}\overline{Q}_i\phi^\dagger u_i + Y^d_{ij}\overline{Q}_i\phi d_i + Y^e_{ij}\overline{L}_i\phi e_i + \text{h.c.}, \tag{2.14}$$

where Y^u, Y^d, and Y^e are 3×3 complex matrices containing dimensionless couplings. After SSB, as described in Section 2.1.1, ϕ acquires a vev so that $\langle\phi\rangle = (0, v/\sqrt{2})$. In effect, the field ϕ mixes the right- and left-handed fermion

fields with terms analogous to the Dirac mass terms in Eq. (2.10). The Yukawa matrices Y^u, Y^d, and Y^e are not diagonal in the $SU(2)$ flavor basis, giving rise to flavor mixing and oscillations in the SM. These effects are discussed further in Section 2.1.2.

2.1.2 Boson and fermion masses

The SM gauge bosons and fermions acquire mass terms by the Higgs mechanism. After SSB, the gauge fields $W_{b\mu}$ and B_μ defined in the SM covariant derivative (Eq. (2.8)) give rise to the gauge boson mass eigenstates W^+, W^-, Z^0, and A. The three weak bosons correspond to the spontaneously broken $SU(2)$ symmetry, and attain nonzero masses. The photon corresponds to the unbroken $U(1)_{\mathrm{EM}}$ gauge symmetry, and remains massless. The W and Z boson masses are proportional to the Higgs vev,

$$\mathrm{M_W} = g\frac{v}{2}, \qquad \mathrm{M_Z} = \sqrt{g^2 + g'^2}\frac{v}{2}, \tag{2.15}$$

where g and g' are the electroweak couplings. It is also convenient to define the weak mixing angle, θ_w, relating these two masses:

$$\mathrm{M_W} = \mathrm{M_Z}\cos\theta_w. \tag{2.16}$$

The weak mixing angle, the couplings g and g', and the electric charge e are all related by:

$$e = \frac{gg'}{\sqrt{g^2 + g'^2}} = g\sin\theta_w. \tag{2.17}$$

This relation reflects change of basis between the gauge bosons W_b^μ and B^μ, and their mass eigenstates W^+, W^-, Z^0, and A.

The Higgs boson corresponds to excitations of the scalar field ϕ in the radial

direction of the complex plane. After SSB, we can write,

$$\phi(x) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + \mathrm{h}(x) \end{pmatrix}, \tag{2.18}$$

where $\mathrm{h}(x)$ is a scalar particle corresponding to the Higgs boson. The mass of this boson is given by:

$$\mathrm{M_h} = \sqrt{2}\mu = \sqrt{2\lambda}v, \tag{2.19}$$

where μ and λ are the coupling constants appearing in the Higgs potential in Eq. (2.12).

The fermion mass terms come from the Yukawa sector of $\mathcal{L}_{\mathrm{SM}}$. Substituting the Higgs vev for ϕ in Eq. (2.14) yields masses of the form,

$$\mathrm{M}_f = -\frac{1}{\sqrt{2}} y_f v, \tag{2.20}$$

where M_f is the fermion's mass, and y_f is its Yukawa coupling, determined by the relevant 3×3 Yukawa matrix defined in Eq. (2.14).

In Run 1 of the LHC, the observation of the Higgs boson with $\mathrm{M_h} \sim 125\,\mathrm{GeV}$ has allowed the Higgs mechanism to be probed experimentally. In particular, the SM predicts that the masses of the fermions and gauge bosons should be proportional to their Higgs couplings. Fig. 2.2 confirms this relationship, where couplings determined by measurements of the Higgs boson production and decay rates are compared with known particle masses.

2.1.3 Quark masses, flavor mixing, and QCD

Quark masses in the SM arise from the Yukawa couplings in Eq. (2.14). The Yukawa matrices Y^u and Y^d are diagonalized by four unitary matrices $V_{L,R}^{u,d}$ to

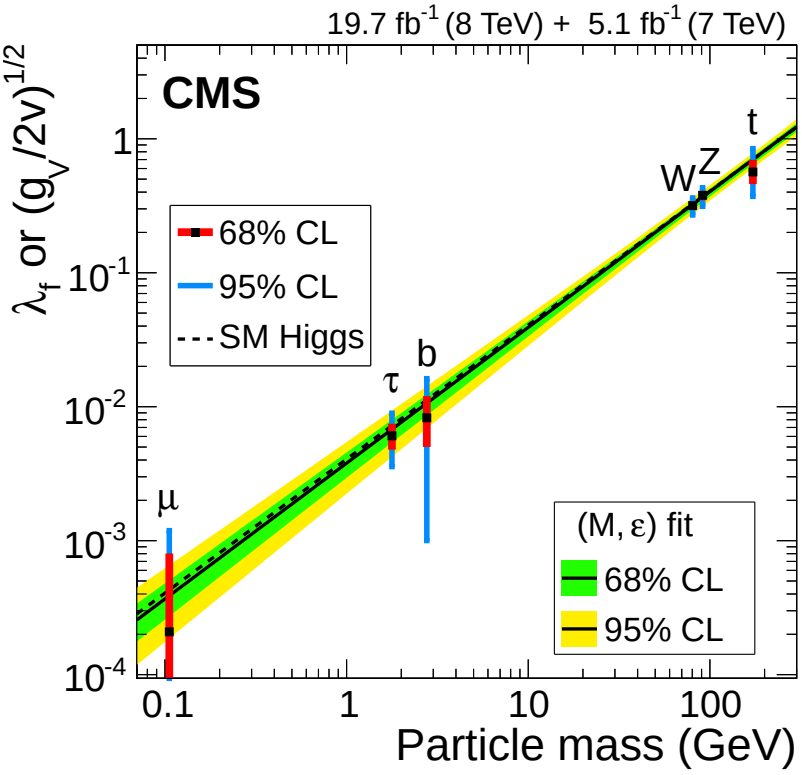

Figure 2.2: Measured couplings of particles to the Higgs boson versus particle masses. The results are consistent with the SM Higgs boson (dashed line) [3].

obtain diagonal mass matrices of the form:

$$\mathrm{M}^{f}_{\mathrm{diag}} = V^{f}_{L} Y^{f} V^{f\dagger}_{R} (v/\sqrt{2}), \tag{2.21}$$

where $f = u, d$. However, in the mass basis, the couplings of the quarks to the charged-current interaction are off-diagonal. The Cabibbo-Kobayashia-Maskawa (CKM) matrix is a 3×3 unitary matrix that contains these couplings [34]. It is given by:

$$V_{\mathrm{CKM}} = V^{u}_{L} V^{d\dagger}_{L} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix}. \tag{2.22}$$

Each element of this matrix is complex-valued, and specifies the coupling between the respective up-type and down-type quarks through the $W^{\pm}$ interactions. It can be parametrized in terms of four parameters – three mixing angles and one complex CP-violating phase. The magnitudes of the CKM matrix elements, $|V_{ij}|$, are determined by measuring quark and meson lifetimes, decay rates, branching fractions, and oscillations. The most precise determination of the CKM matrix element magnitudes comes from global fits incorporating all available measurements and imposing SM constraints. A recent fit gives the magnitudes [34]:

$$V_{\mathrm{CKM}} = \begin{pmatrix} 0.97428 \pm 0.00015 & 0.2253 \pm 0.0007 & 0.0347^{+0.00016}_{-0.00012} \\ 0.2252 \pm 0.0007 & 0.97345^{+0.00015}_{-0.00016} & 0.0410^{+0.0011}_{-0.0007} \\ 0.00862^{+0.00026}_{-0.00020} & 0.0403^{+0.0011}_{-0.0007} & 0.999152^{+0.000030}_{-0.000045} \end{pmatrix}. \quad (2.23)$$

The complex phases are constrained in measurements of CP violation in meson mixing and decays. It is evident from Eq. (2.23) that the CKM matrix in the SM is diagonal to first order. The top quark, discussed in Section 2.2, has $|V_{\mathrm{tb}}| \approx 1$, forcing it to decay almost exclusively through the mode $\mathrm{t} \to \mathrm{Wb}$. Conversely, the bottom quark cannot decay to the top quark due to kinematic constraints, and is forced to decay through channels with relatively low values of $|V_{\mathrm{ub}}|$ and $|V_{\mathrm{cb}}|$. For this reason, despite the bottom quark's large mass ($\mathrm{M_b} \approx 4.2\,\mathrm{GeV}$) in relation to its decay products, bottom hadrons have large lifetimes of $\sim 10^{-12}$ s.

The values of the quark masses in Eq. (2.21) are determined by a combination of experimental and theoretical techniques [34]. The challenge in measuring these parameters stems from the confinement of quarks into hadrons, and the nonperturbative nature of quantum chromodynamics (QCD) in the low energy regime below $\sim 1\,\mathrm{GeV}$.

The light quarks – up, down, and strange – have masses below 1 GeV. The most reliable determinations of the masses $\mathrm{M_s}$ and $\mathrm{M_{ud}} = (\mathrm{M_u} + \mathrm{M_d})/2$ is given

by lattice gauge theory [34], which provides a simulation of QCD that does not rely on a perturbative expansion in the strong coupling constant. Experimental measurements of meson masses are input into the simulation to back out the values of the bare quark masses. In addition to lattice QCD, ratios of light quark masses can be obtained using chiral perturbation theory.

Heavy quark masses satisfy $Q \gg \Lambda_{QCD}$, and can be extracted using heavy quark effective field theory. Here, mesons containing one bottom or charm quark and one light quark are treated as a two-quark bound state. In this bound state, the mass of the 'nucleus', i.e. the heavy quark, can be related to the total mass of the system. Alternatively, mesons containing two heavy quarks, such as the J/ψ or Υ, can be approximated as bound states in non relativistic QCD, where the calculation includes an expansion in $\beta = v/c$. Of all the quarks, the top quark has the shortest lifetime of $\sim 5 \times 10^{-25}$ s, and decays before it can hadronize. For this reason, its mass can be measured directly by a reconstruction of its decay products. In interpreting measurements of the top quark mass, complications arise due to effects in nonperturbative QCD – a further discussion on this topic is given in Section 2.2.

2.2 The top quark

The top quark is contained in the third generation of the Q and d_R fields specified in Table 2.1. It was postulated by Kobayashi and Maskawa in 1973 along with the bottom quark [35], and discovered in 1995 by the CDF [36, 37] and D0 [38] experiments at Fermilab. It is a spin-$\frac{1}{2}$ fermion with an electric charge of $\frac{2}{3}e$, a large mass of $\sim 173\,\text{GeV}$, and a short lifetime of 5×10^{-25} s. The term of $\mathcal{L}_{\text{Yuk}}$

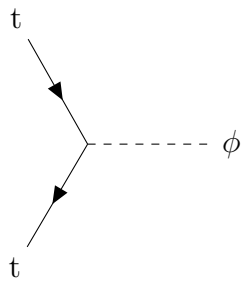

Figure 2.3: Top-Higgs coupling in the SM, described by Eq. (2.24).

corresponding to the top quark is given by,

$$\Delta\mathcal{L} = -y_{\mathrm{t}}\bar{\mathrm{t}}_L\phi\mathrm{t}_R, \tag{2.24}$$

with the interaction vertex shown in Fig. 2.3. After SSB, this coupling generates a mass term of the form in Eq. (2.20). The top quark's Yukawa coupling, responsible for its large mass, has a value of $y_{\mathrm{t}} = \sqrt{2}\mathrm{M}_{\mathrm{t}}/v \approx 1$.

In the electroweak sector, the top quark interactions after SSB are contained in the terms:

$$\begin{aligned}\Delta\mathcal{L} = &-\frac{g}{2\sqrt{2}}(J^{\mu}_{\mathrm{W}}\mathrm{W}^{-}_{\mu} + J^{\mu\dagger}_{\mathrm{W}}\mathrm{W}^{+}_{\mu}) \\ &-\frac{\sqrt{g^2+g'^2}}{2}J^{\mu}_{\mathrm{Z}}\mathrm{Z}_{\mu} - \frac{gg'}{\sqrt{g^2+g'^2}}J^{\mu}_{Q}\mathrm{A}_{\mu},\end{aligned} \tag{2.25}$$

where the top quark contributions to the currents J^{μ}_{W}, J^{μ}_{Z}, J^{μ}_{Q} are given by:

$$\Delta J^{\mu\dagger}_{\mathrm{W}} = 2\bar{\mathrm{b}}_L\gamma^{\mu}V_{\mathrm{CKM}}\mathrm{t}_L \tag{2.26}$$

$$\Delta J^{\mu}_{\mathrm{W}} = \bar{\mathrm{t}}_L\gamma^{\mu}V_{\mathrm{CKM}}\mathrm{b}_L \tag{2.27}$$

$$\Delta J^{\mu}_{Q} = -\frac{1}{3}\bar{\mathrm{t}}_L\gamma^{\mu}\mathrm{t}_L - \frac{1}{3}\bar{\mathrm{t}}_R\gamma^{\mu}\mathrm{t}_R \tag{2.28}$$

$$\Delta J^{\mu}_{\mathrm{Z}} = -\bar{\mathrm{t}}_L\gamma^{\mu}\mathrm{t}_L - 2\sin^2\theta_W J^{\mu}_{Q}. \tag{2.29}$$

In the charged-current interaction, we have included the CKM matrix V_{CKM}, which reflects the change of basis between the mass and flavor eigenstates of the quarks.

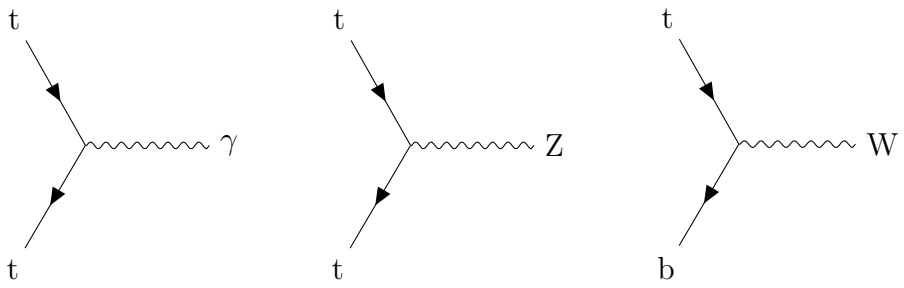

Figure 2.4: Top electroweak couplings in the SM, described by Eq. (2.25).

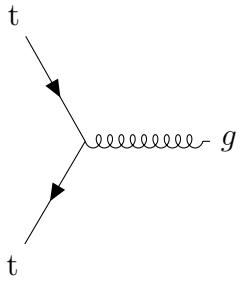

Figure 2.5: Top-gluon coupling in the SM, described by Eq. (2.30).

For the top quark, the mixing provided by the CKM matrix is minimal. The vertices corresponding to these terms are shown in Fig. 2.4.

In the QCD sector, the top-gluon coupling terms can be written:

$$\Delta\mathcal{L} = -g_s\bar{\mathrm{t}}_L\lambda_a\gamma_\mu G_a^\mu \mathrm{t}_L - g_s\bar{\mathrm{t}}_R\lambda_a\gamma_\mu G_a^\mu \mathrm{t}_R, \tag{2.30}$$

where the index a runs over eight gluon fields. The corresponding interaction vertex is shown in Fig. 2.5.

2.2.1 Production and decay

At the LHC, top quarks are produced predominantly in pairs by the strong-interaction processes shown in Fig. 2.6. In proton-proton collisions at $\sqrt{s} = 8$ TeV,

gluon-gluon fusion accounts for approximately 85% of $t\bar{t}$ production, with quark-quark annihilation responsible for the remaining 15%. Measurements of the top quark production cross section provide important tests of QCD and electroweak theory, and can be sensitive to BSM physics. As of 2016, the $t\bar{t}$ cross section has been measured at five center-of-mass energies, including $\sqrt{s} = 1.8$, 1.96 TeV at the Tevatron, and $\sqrt{s} = 7$, 8, 13 TeV at the LHC. The total theoretical cross section for $t\bar{t}$ production at $\sqrt{s} = 8\,\text{TeV}$ is $\sigma_{t\bar{t}} = 247.7^{+6.3+11.5}_{-8.5-11.5}$ pb [39] assuming $M_t = 173.3\,\text{GeV}$. The dependence of the $t\bar{t}$ production cross section on the center-of-mass energy is shown in Fig. 2.7, where theoretical predictions are compared to cross section measurements at the Tevatron and LHC.

The differential $t\bar{t}$ cross section has been measured as a function of variables such as the $t\bar{t}$ invariant mass, p_T, and η, as well as the p_T and η of the t and $\bar{t}$. Recent measurements of the top quark p_T spectrum have revealed minor disagreements between data and theoretical predictions at leading order (LO), where the data is observed to have a softer spectrum at large values of the $t\bar{t}$ invariant mass and in the tail of the top quark p_T spectrum [40, 41]. At next-to-leading-order (NLO) in the theory, the agreement in these spectra is improved. Measurements of the top quark mass at CMS include a systematic uncertainty to account for these effects.

At the LHC, single top quarks are produced by the electroweak processes shown in Fig. 2.8. These production mechanisms occur at a lower cross section compared to $t\bar{t}$ production, and are more difficult to probe due to lower statistics and large background contributions from $t\bar{t}$ production. The t-channel production has the largest $t + \bar{t}$ production cross section, predicted to be $\sigma_{\text{t-ch}} = 87.1^{+0.24}_{-0.24}$ pb at $\sqrt{s} = 8\,\text{TeV}$ with $M_t = 173.3\,\text{GeV}$. The relative proportion of t and $\bar{t}$ production is

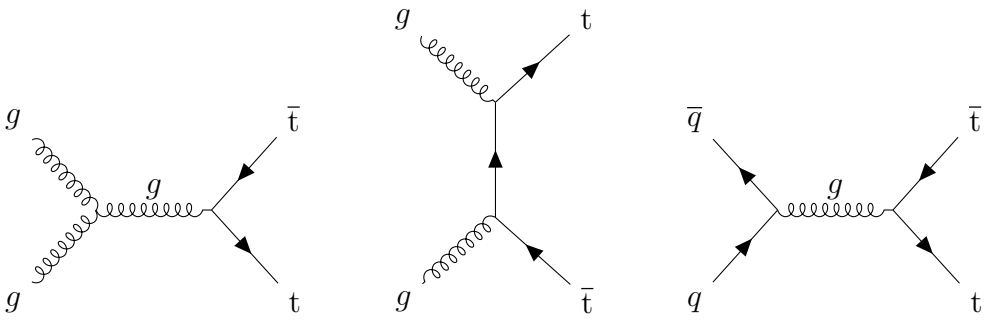

Figure 2.6: Diagrams contributing to $t\bar{t}$ production at the LHC. They include (left) gluon-gluon fusion, the highest-cross-section process, (center) t-channel production, and (right) quark-antiquark annihilation.

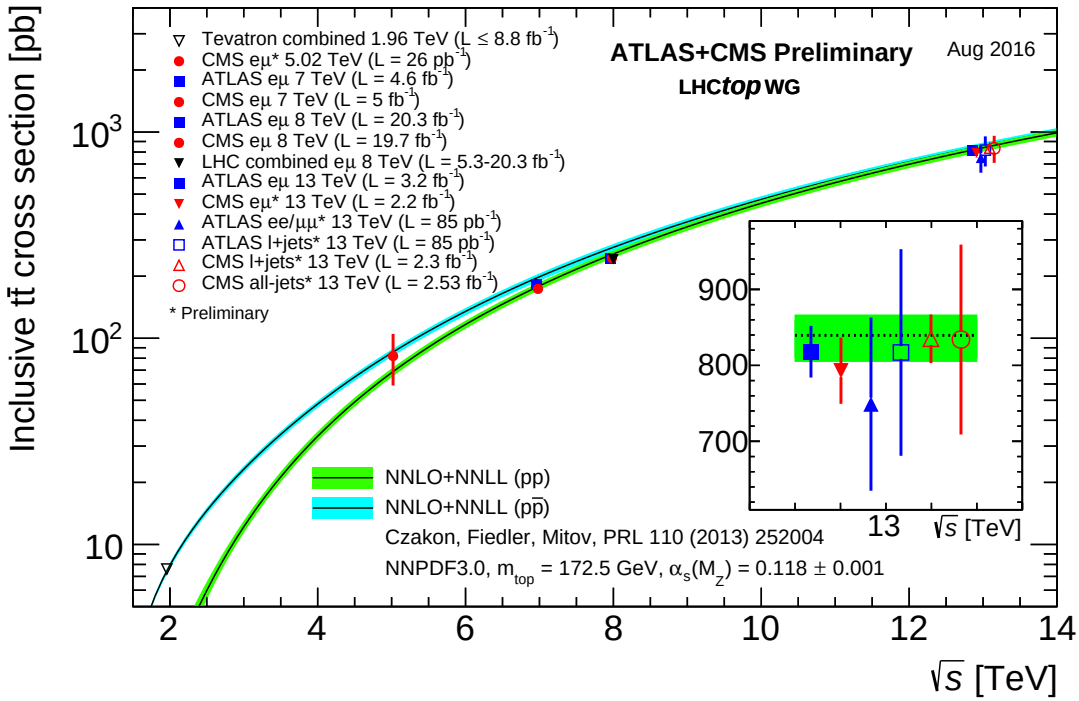

Figure 2.7: Summary of LHC and Tevatron $t\bar{t}$ cross section measurements, compared to a QCD NNLO calculation complemented with NNLL resummation (top++2.0). All cross sections assume $M_t = 172.5\,\text{GeV}$ [4].

65% and 35%, respectively [42]. The theoretical predictions for the s-channel and tW-channel cross sections are $\sigma_{\text{s-ch}} = 4.6 \pm 0.3\text{pb}$ [43] and $\sigma_{\text{tW-ch}} = 15.6 \pm 1.2\text{pb}$ [44], respectively. The cross section values for single top production are an important probe for BSM physics. Single top measurements also provide constraints on the value of the $|V_{\text{tb}}|$ CKM matrix element [34].

The decay of the top-quark is expected to be dominated by the $\text{t} \to \text{Wb}$ channel, where $|V_{\text{tb}}| \gg |V_{\text{td}}|, |V_{\text{ts}}|$ can be assumed. The top-quark decay width, as predicted by the SM at NLO, can be written,

$$\Gamma_{\text{t}} = \frac{G_F \text{M}_{\text{t}}^3}{8\pi\sqrt{2}} \left(1 - \frac{\text{M}_{\text{W}}^2}{\text{M}_{\text{t}}^2}\right)^2 \left(1 + 2\frac{\text{M}_{\text{W}}^2}{\text{M}_{\text{t}}^2}\right) \left[1 - \frac{2\alpha_s}{3\pi}\left(\frac{2\pi^2}{3} - \frac{5}{2}\right)\right], \tag{2.31}$$

where M_{t} in this expression represents the top quark pole mass [34]. For $\text{M}_{\text{t}} = 173.3\,\text{GeV}$, this relation predicts a width of $\Gamma_{\text{t}} = 1.35\,\text{GeV}$, which corresponds to a lifetime of 5×10^{-25} s. With this lifetime, the top quark decay occurs an order of magnitude faster than the timescale for hadronization to occur. For this reason, the top quark does not form top-flavored hadrons or $\text{t}\bar{\text{t}}$ bound states.

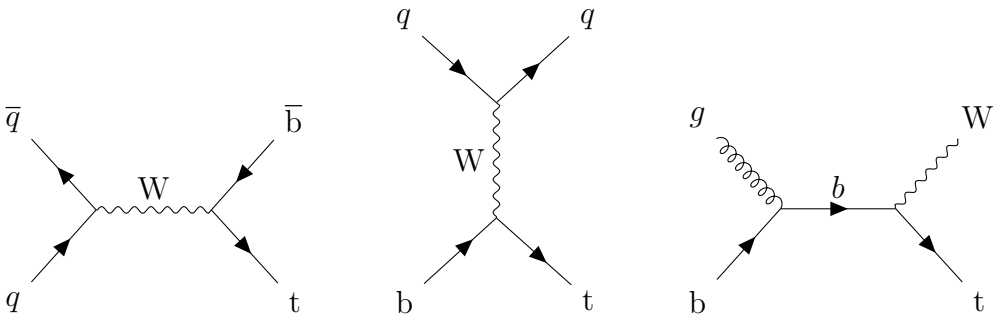

Figure 2.8: Diagrams contributing to single-top production at the LHC. They include (left) s-channel, (center) t-channel, and (right) tW-channel production.

The $\text{t}\bar{\text{t}}$ decay produces a two-step topology, with $\text{t} \to \text{Wb}$ in the first step, followed by $\text{W} \to jj$ or $\text{W} \to \ell\nu$ in the second. Given this topology, $\text{t}\bar{\text{t}}$ events can be classified into three decay channels determined by the two W boson decays.

The all-hadronic decay channel contains two hadronically-decaying W bosons, and has a branching fraction of 45.7%; the semileptonic channel has one hadronic and one leptonic W boson decay, with a branching fraction of 43.8%; lastly, the dileptonic channel is the one where both W bosons decay leptonically, with a branching fraction of 10.5%. The branching fractions of these decay modes are graphically illustrated in Fig. 2.9. Each of these channels has particular strengths and weaknesses when employed for top quark mass reconstruction. The dileptonic channel, featured in Chapter 7, provides a clean dilepton event selection, but is kinematically underconstrained due to the production of two neutrinos – it is impossible to construct the invariant masses $M_{b\ell\nu}$ for both top quarks in a single event. The all-hadronic channel has the benefit of large statistics, but also contains significant backgrounds due to QCD multijet production. The semileptonic channel provides the best overall precision for measurements of the top quark mass, with a balance between a clean event selection, good statistics, and kinematics that are well-suited for mass reconstruction.

2.2.2 Top quark mass

The top quark mass is a free parameter of the SM, and its value is an important input for precision tests of the SM [6, 45]. In the 1990s, experiments at the Large Electron Positron (LEP) collider, Stanford Linear Collider (SLC), and Tevatron provided measurements of a number of SM observables, including the masses, widths, and decays of the electroweak bosons. The precision achieved by these experiments is sufficient to probe quantum loop corrections to these observables stemming from the SM fermions and the Higgs. In particular, the top quark mass plays a relatively large role in these radiative corrections due to its large difference

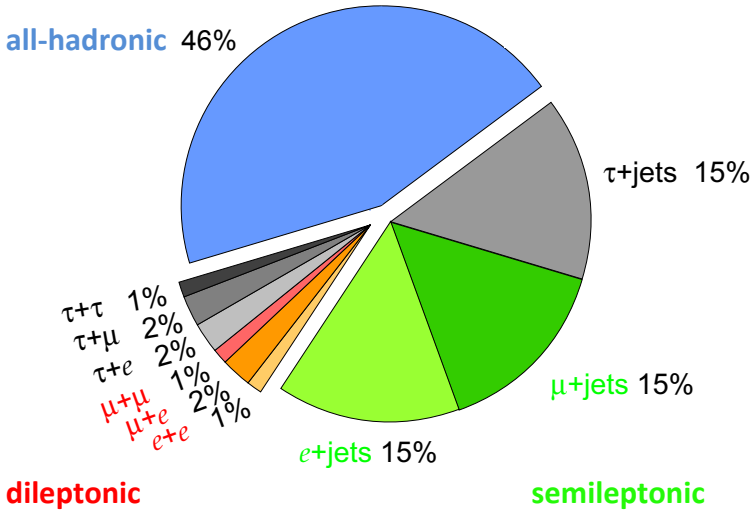

Figure 2.9: Branching fractions of the $t\bar{t}$ decay modes, modified from [5].

with the mass of its isospin partner, the bottom quark. The value of M_t obtained in direct measurements, when compared to the value predicted by global fits to electroweak data, provides a test of the self-consistency of the SM.

A detailed review of electroweak observables and their dependence on the value of M_t is given in Ref. [45]. The relevant observables include: the Z boson mass, width, and partial decay widths to hadrons and leptons; the Z asymmetries in leptonic and hadronic decay modes; the value of the weak mixing angle in Eq. (2.16) as determined by Z decays; and, the W boson mass. These observables can be combined into a global fit to predict the values of M_t and M_h, or if these values are known from direct measurements, their agreement with the theoretically-preferred values can be tested. Examples of processes that play an important role in global electroweak fits are shown in Figs. 2.10 and 2.11 – they include quantum loop corrections from the top quark and Higgs boson to the W boson mass, as well as top quark and W boson contributions to the $Z \to b\bar{b}$ decay width.

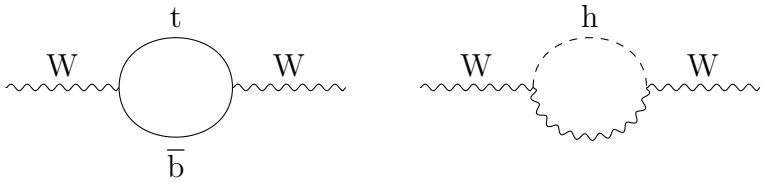

Figure 2.10: Lowest-order radiative corrections to the W boson mass involving (left) top and bottom quarks, and (right) the Higgs boson.

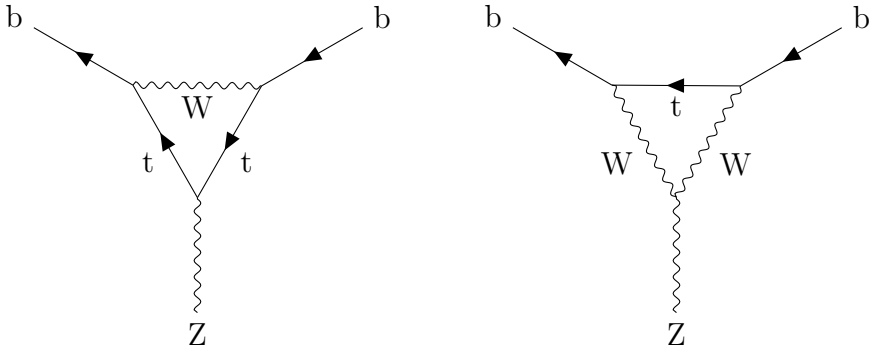

Figure 2.11: Quantum loop corrections involving the top quark make significant contributions to the width of the $Z \to b\bar{b}$ decay process.

In a recent global electroweak fit, an agreement of $p = 0.22$ is found between SM predictions and direct measurements [6]. In this fit, the value of M_t determined in the fit agrees to within $0.5\sigma_{exp}$ with direct measurements, where σ_{exp} is the experimental uncertainty on M_t. The measured values of M_t and M_W are compared with theoretical predictions in Fig. 2.12. In the fit, M_t and M_W are observed to be consistent with the measured Higgs boson mass of $\sim 125\,\text{GeV}$.

The top quark Yukawa coupling also has implications for the shape of the Higgs potential, given in Eq. (2.12), through its contribution to the Higgs quartic coupling, λ. For the Higgs potential to retain its 'Mexican hat' shape, we must have $\mu^2 > 0$ and $\lambda > 0$. At the electroweak scale, the values are indeed positive, with $\mu^2 \approx (88\,\text{GeV})^2$ and $\lambda \approx 0.13$. However, when evolved to higher energy scales, quantum loop corrections from the Yukawa sector contribute to a decreasing value

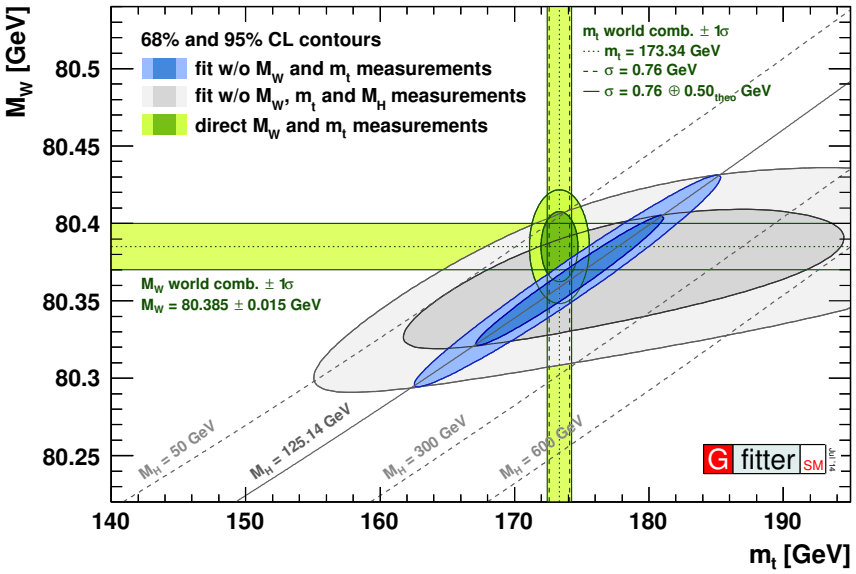

Figure 2.12: Confidence intervals for M_W versus M_t in the fit, with and without including the Higgs mass constraint. The central value is consistent with direct measurements of these masses. This fit has a p-value of 0.22 for consistency between the SM and direct measurements [6].

of λ. The largest Yukawa coupling comes from the top quark, with $y_t \sim 1$. Fig. 2.13 shows the relevant parameters space, defined by values of the Higgs and top quark masses, where the evolution of λ has been computed to NNLO [7]. In the stable region, the Higgs couplings satisfy $\mu^2 > 0$ and $\lambda > 0$. In the other regions, the value of λ may become negative at energy scales above 10^9 GeV, forming a second minimum in the Higgs potential. If the probability of tunneling to the second minimum is small on universal timescales, the region is meta-stable. Large tunneling probabilities correspond to unstable regions of the parameter space. The currently known values of M_t and M_h point to a region of the parameter space that is meta-stable. It is important to note that the evolution of λ computed in this study assumes that no BSM physics exists up to high energy scales. For this reason, the results must be interpreted with care. Nevertheless, the measured values of

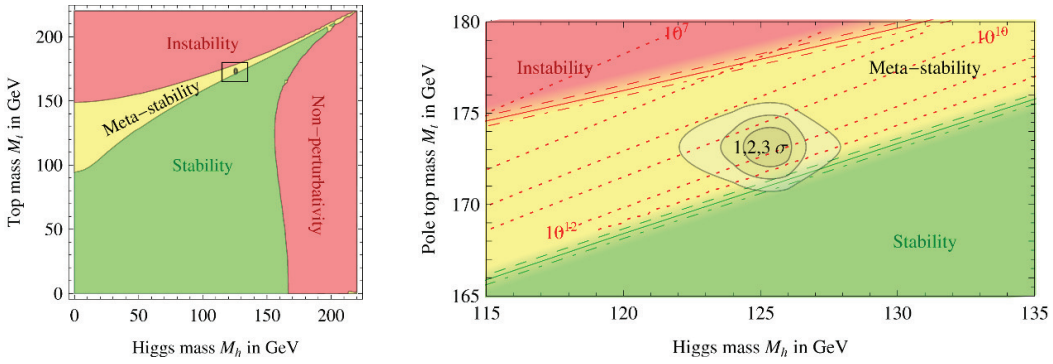

Figure 2.13: Regions of stability, meta-stability, and instability of the SM vacuum are shown in green, yellow, and red, respectively, in the M_t-M_H plane. Direct measurements of these masses occupy the region shown in gray (right) [7].

M_t and M_h seem to occupy a special region of the parameter space, characterized by two minima in the SM Higgs potential – one at the electroweak scale, and the other near the Planck scale. Whether this is a mere coincidence or something of physical significance remains to be investigated.

At this stage, it is important to bring attention to the definition and interpretation of M_t, the mass of an object carrying a nonzero QCD color charge, which is subject to nonperturbative infrared effects in QCD [8, 46, 47]. The mass of a directly-observable particle, such as the electron, is well-defined as the position of the pole in electron's propagator. For quarks, this definition, known as the 'pole mass', is valid within the framework of perturbation theory. The propagator of a quark with four-momentum p can be written:

$$D(\not{p}) = \frac{i}{\not{p} - M_R - \Sigma(\not{p})}, \tag{2.32}$$

where M_R is a renormalized short-distance mass, such as the $\overline{\text{MS}}$ mass, evaluated at a scale $\mu \gg \Lambda_{\text{QCD}}$, and $\Sigma(\not{p})$ is the renormalized quark self energy [8]. The

location of the pole in this propagator is given by:

$$\not{p}_{\text{pole}} = \mathrm{M}_R + \Sigma(\not{p}_{\text{pole}}). \tag{2.33}$$

Although M_R is well-defined in this expression, the self energy $\Sigma(\not{p}_{\text{pole}})$ cannot be evaluated to arbitrary precision in perturbation theory. As a result, the pole mass of a quark is ambiguous by an amount of order $\Lambda_{\text{QCD}} \approx 200\,\text{MeV}$.

The source of the ambiguity in the top quark pole mass can be understood by considering the decay $\mathrm{t} \to \mathrm{Wb}$. In a perturbative treatment of the decay, as in Fig. 2.14(a), the top quark pole mass can be extracted by simply reconstructing the Wb invariant mass. A full treatment of this process, however, must account for the neutralization of the top quark's color as it decays to the color-neutral W boson and b jet. The colorless hadrons that make up the b jet must contain at least one quark external to the $\mathrm{t} \to \mathrm{Wb}$ decay, as shown in Fig. 2.14(b). The momentum of this additional quark is of order Λ_{QCD}, and contributes to the ambiguity in the top quark pole mass.

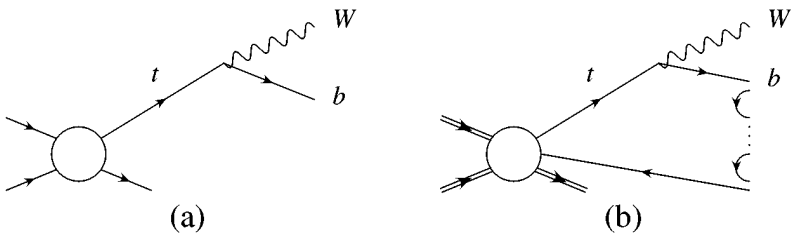

Figure 2.14: The production and decay of a top quark in (a) perturbation theory, and (b) including nonperturbative effects [8].

Typical measurements of $\mathrm{M_t}$ are calibrated to the definition of the top quark mass in MC simulation, known as the 'MC mass', or $\mathrm{M_t^{MC}}$. A sketch of a single proton-proton collision, as simulated by the SHERPA event generator, is shown in Fig. 2.15. Nonperturbative aspects of the simulation, such hadronization and

parton showering, are parametrized and then tuned to match observed data. As a result, the connection between M_t^{MC} and a well-defined mass, such as the top quark pole mass or $\overline{MS}$ mass, is estimated to carry an uncertainty of $\sim 1\,\text{GeV}$ [47].

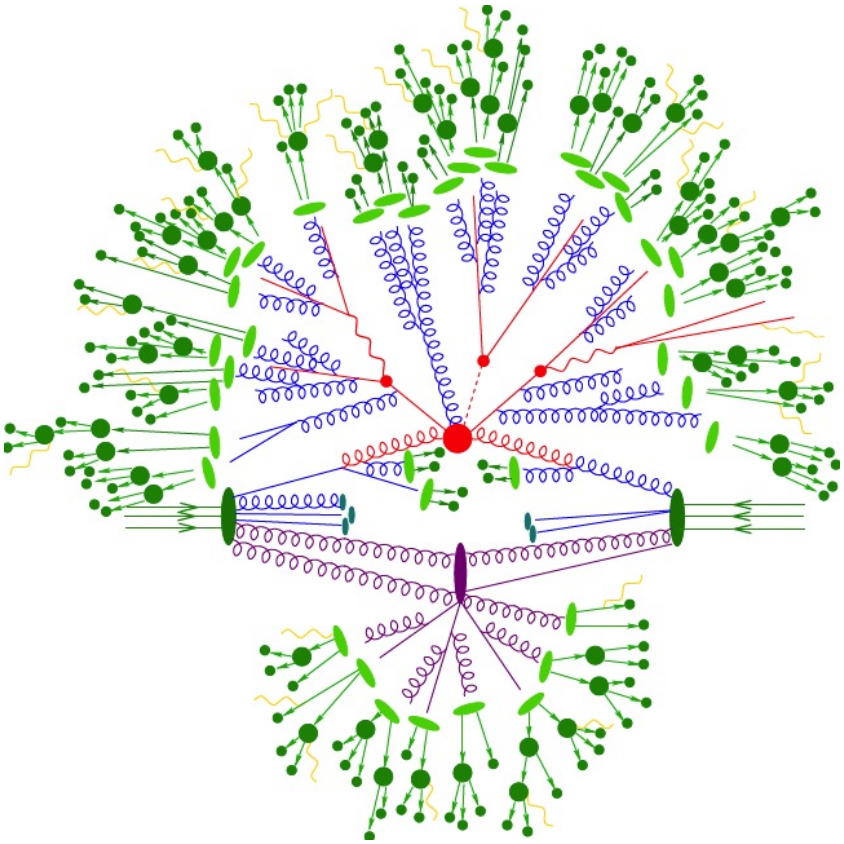

Figure 2.15: Sketch of a proton-proton collision in the SHERPA event generator. Shown are the matrix-element calculations (red), initial-state parton showers (blue), multiple-parton interactions and underlying event (purple), fragmentation (light green), hadron decay (dark green), and final-state radiation (yellow) [9].

Alternative methods attempt to determine the value of M_t without relying on a calibration to MC simulation. The value of the top quark pole mass can be directly extracted from the inclusive $t\bar{t}$ production cross section to a precision of approximately $\pm 3\,\text{GeV}$ [48]. Although this method is directly sensitive to the pole mass, its precision is somewhat limited compared to standard methods. The top quark mass has also been determined using the endpoints of kinematic dis-

tributions, as documented in Ref. [29], a measurement involving the author. The endpoint method does not rely on template shapes determined in MC simulation, and therefore does not conduct a direct calibration to M_t^{MC}. It provides a determination of M_t that is minimally-reliant on MC simulation. The measurement outlined in Chapter 7 does require a calibration to MC simulation, and carries the corresponding theoretical ambiguity of $\sim 1\,\text{GeV}$.

CHAPTER 3

EXPERIMENTAL APPARATUS

We're machines for turning caffeine into physics.

– Nima Arkani-Hamed

The Compact Muon Solenoid (CMS) is a general-purpose detector positioned on the Large Hadron Collider (LHC). The broad experimental program at CMS includes SM measurements in the Higgs, top quark, QCD, and electroweak sectors, searches for BSM physics with a range of potential signatures, and the study of heavy ion collisions. In July of 2012, the CMS collaboration announced the co-discovery of the Higgs boson [49], confirming the role of the Higgs in electroweak symmetry breaking and setting a milestone in the study of TeV-scale physics. In this chapter, we give an overview of the LHC, as well as CMS and its subdetector systems.

3.1 The Large Hadron Collider

The LHC is a proton-proton collider at CERN near Geneva, Switzerland [1]. With a design proton-proton center-of-mass energy of 14 TeV, it is the highest-energy collider ever built. The LHC storage ring is housed in a tunnel of 27 km circumference, located about 100 m beneath the France-Switzerland border. Proton beams collide at four points along the storage ring, providing collisions to the ATLAS, CMS, ALICE, and LHCb experiments.

The LHC began stable operation with $\sqrt{s} = 7\,\text{TeV}$ in 2010, after a two-year delay due to a magnet quench incident. The center-of-mass energy was increased to $\sqrt{s} = 8\,\text{TeV}$ in 2012. After a long shutdown in 2013-2014, the energy was increased to $\sqrt{s} = 13\,\text{TeV}$ for data taking in 2015 and 2016. From the beginning of its proton-proton operation in 2010 to the end of 2016, the LHC has delivered an integrated luminosity corresponding to approximately $75\,\text{fb}^{-1}$ to its experiments.

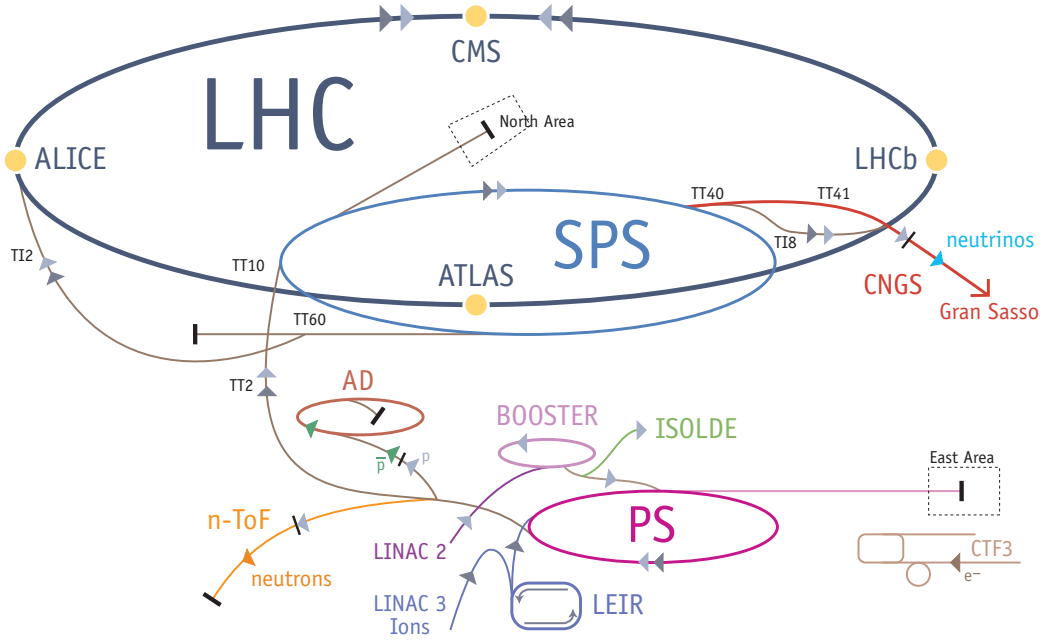

Figure 3.1: The CERN accelerator complex and the LHC [10].

Before reaching its nominal energy in the LHC storage ring, each beam proton traverses a series of pre-accelerators, shown in Fig. 3.1. Protons are produced by stripping the electrons from hydrogen gas, and are injected into Linac 2, where they reach an energy of 50 MeV. The protons are then accelerated up to 1.4 GeV in the Proton Synchrotron Booster (PSB), up to 25 GeV in the Proton Synchrotron (PS), and up to 450 GeV in the Super Proton Synchrotron (SPS). In the final step of the acceleration sequence, protons are injected into the LHC in clockwise and

counter-clockwise directions, where they are accelerated to their nominal energy of up to 7 TeV.

The LHC beams are controlled by a series of magnets and radio-frequency (RF) cavities arranged into eight sectors around the storage ring, shown in Fig. 3.2. The LHC employs 1232 dipole magnets, which provide an 8 T bending magnetic field to guide the proton beams in their near-circular orbit around the ring. They are constructed using superconducting NbTi cables, which are bathed in liquid helium at a temperature of 1.9 K. The proton beams are focused by 392 quadrapole magnets placed along the ring, and further corrected by magnets of higher multipole orders. The bunch structure and acceleration of each beam from 450 GeV to its nominal energy is controlled by eight superconducting RF cavities. Each cavity provides an accelerating field of 5 MV/m with an oscillation frequency of 400 MHz. The LHC proton beams nominally contain about 2800 bunches with a spacing of 25 ns.

Table 3.1: LHC beam parameter values in design conditions [1].

Parameter	Value
E_{p}	7 TeV
N_p	1.15×10^{11}
n_b	2808
f_{rev}	11.245 kHz
γ_r	7461
ϵ_n	3.75 μm rad
σ^*	16.7 μm
β^*	0.55 m
F	0.836

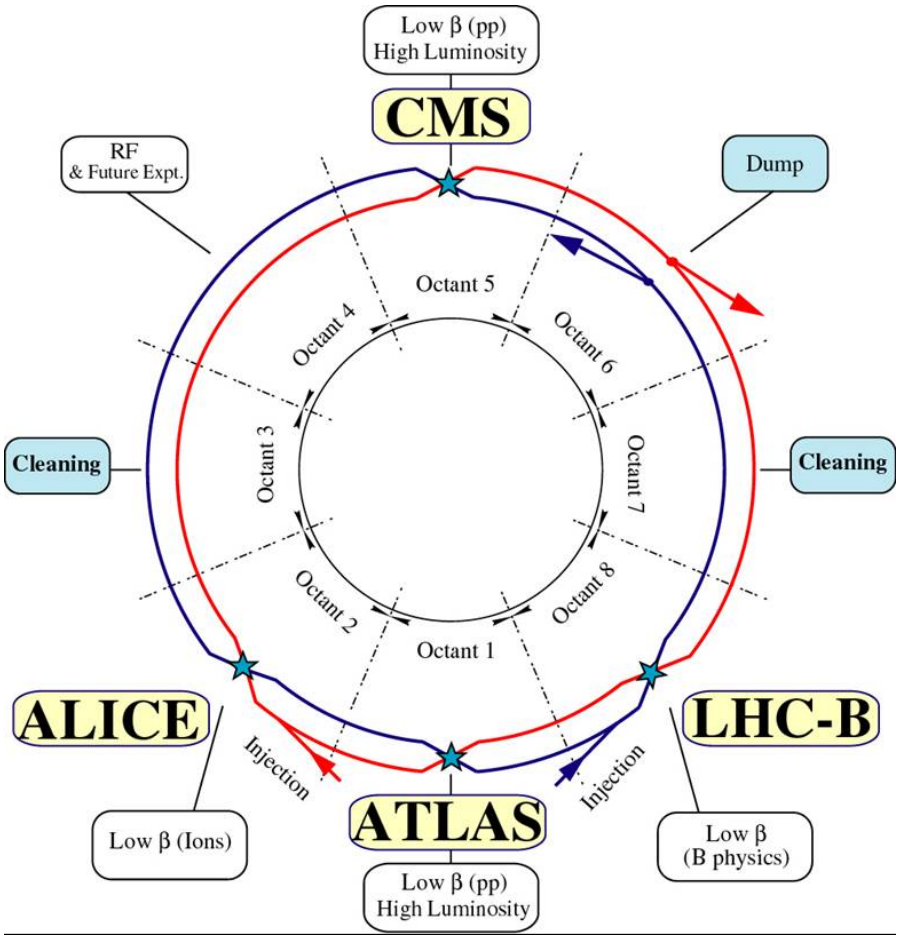

Figure 3.2: Schematic layout of the LHC. Two proton beams, shown in blue and red, are injected into the storage ring in opposite directions. The storage ring contains eight arcs and straight sections. The straight sections are approximately 528m long, and serve as experimental or utility sections. The ring has a total circumference of 27 km [11].

For experiments positioned on the LHC, the event rate, R, is given by the interaction cross section, σ_{int}, and the instantaneous luminosity $\mathcal{L}$:

$$R = \mathcal{L}\sigma_{\text{int}}. \tag{3.1}$$

The instantaneous luminosity is related to the integrated luminosity, $L_{\text{int}} = \int \mathcal{L} \mathrm{d}t$, where the value of $\mathcal{L}$ in a particular LHC run depends on a number of beam

parameters. Assuming a Gaussian transverse beam profile, $\mathcal{L}$ can be written [1]:

$$\mathcal{L} = \frac{N_p^2 n_b f_{\text{rev}} \gamma_r}{4\pi \epsilon_n \beta^*} F, \tag{3.2}$$

where N_p is the number of protons per bunch, limited by the allowable transverse beam size of 1.2 mm; n_b is the number of bunches per beam; f_{rev} is the beam revolution frequency; γ_r is the relativistic gamma factor for each beam; and F is the geometric luminosity reduction factor due to the crossing angle at the interaction point (IP). The transverse beam size is described by $\epsilon_n = \epsilon \beta_r \gamma_r$, the normalized transverse beam emittance, and β^*, the beta function at the IP. The width of the transverse beam profile at the IP, σ^*, is given by:

$$\pi (\sigma^*)^2 = \epsilon \beta^*. \tag{3.3}$$

The transverse emittance, ϵ, is a beam quality characteristic that is determined by the process of bunch preparation, extending back to the proton source [34]. The value of β^* is a feature of the beam optics that is determined by the magnet configuration near the IP. A small value of β^* is desirable to achieve a focused beam at the IP, with a small profile σ^*. The design values of the LHC beam parameters are given in Table 3.1.

3.2 The CMS detector

The CMS detector [21] has a cylindrical geometry, with subdetectors arranged in layers of increasing radius around the IP. A schematic of the detector is shown in Fig. 3.3. A central feature of CMS is its superconducting solenoid, which provides a uniform magnetic field of 3.8 T for the precise measurement of charged particle momenta. The inner tracker is the subdetector closest to the IP. It is an entirely

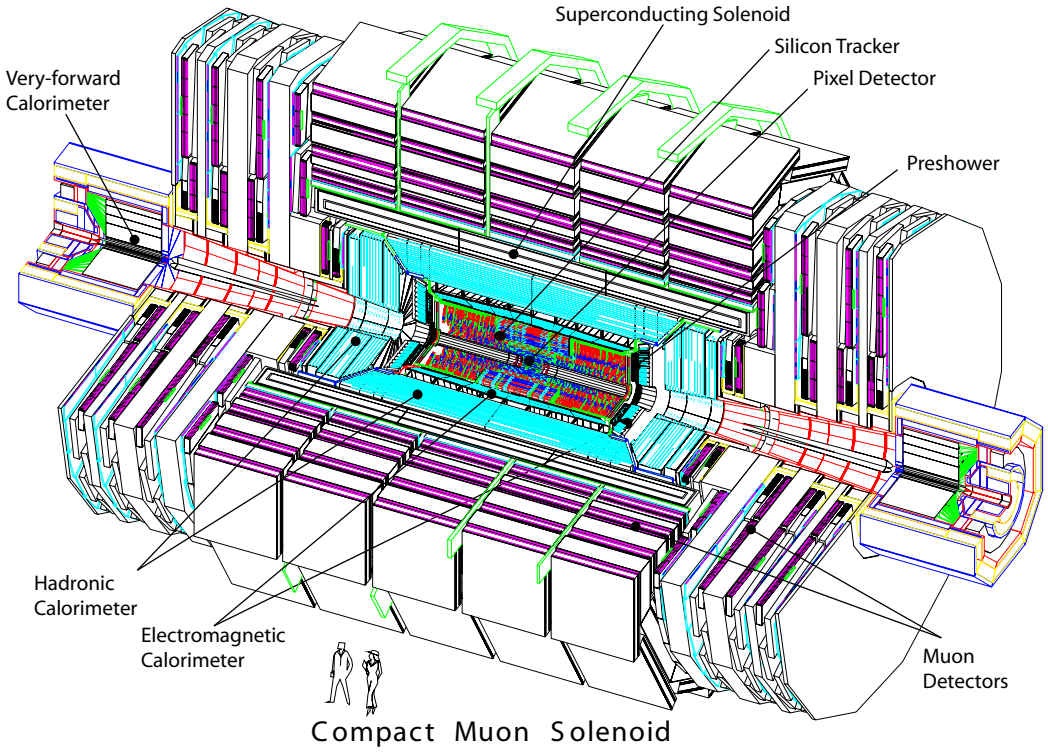

Figure 3.3: Schematic of the CMS detector [12].

silicon-based detector for the reconstruction of charged particle tracks, as well as primary and secondary vertices, inside the tracking volume. The electromagnetic calorimeter (ECAL) positioned outside the inner tracker employs scintillating lead tungstate ($PbWO_4$) crystals to detect and reconstruct electrons and photons. The hadronic calorimeter (HCAL), contained mostly inside the solenoid magnet, is a brass/scintillator sampling calorimeter that measures the energies of hadronic particles that penetrate the ECAL. A closer look at the subdetectors located inside the solenoid coils is given in Fig. 3.4. Outside the solenoid, CMS contains a muon system for precise muon momentum reconstruction and triggering. Overall, the CMS detector is 21.6 m long, has a diameter of 14.6 m, and weighs 12500 tons.

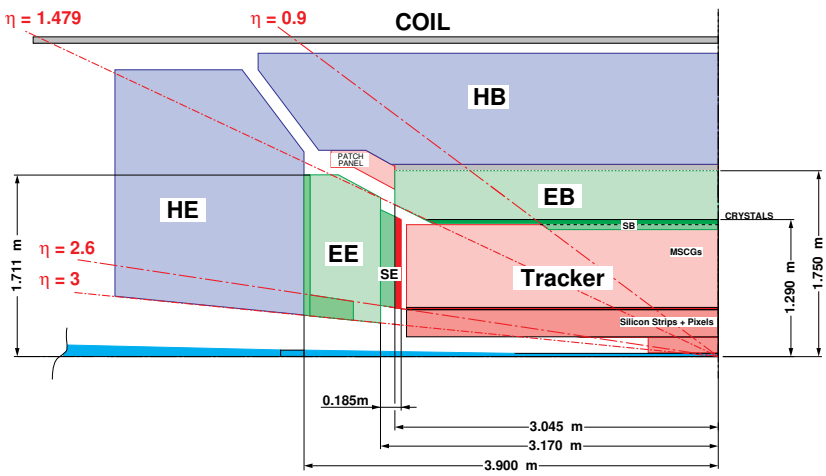

Figure 3.4: Cross section of inner tracker, electromagnetic calorimeter, hadronic calorimeter, and magnet coils [13].

3.2.1 Coordinate system

The CMS detector employs a right-handed coordinate system with the origin corresponding to the IP. In the Cartesian version of this coordinate system, the x-axis points radially inwards towards the center of the LHC ring; the y-axis points vertically upwards; and the z axis points along the LHC beam in a counter-clockwise direction when viewed from above. In addition, polar coordinates are often convenient for describing object trajectories in CMS. Here, the radial coordinate, denoted r, describes the distance from the beam axis; the polar angle θ is measured from the z-axis with $\theta = 90°$ pointing vertically upwards; and the azimuthal angle ϕ moves clockwise in the plane transverse to the beam axis, with $\phi = 0°$ pointing towards the center of the LHC and $\phi = 90°$ pointing vertically upwards.

In LHC collisions, the center-of-mass frame of a hard interaction may be arbitrarily boosted in the z-direction with respect to the lab frame. This is a consequence of the non-pointlike nature of hadrons at sufficiently high energy scales. It is convenient to define a pseudo-rapidity coordinate, η, which is typically substituted

for the polar angle, θ, at CMS. The pseudo-rapidity is a low-mass, high-momentum approximation of the more general rapidity coordinate, y, given by [34]:

$$y = \frac{1}{2} \log \left(\frac{E + p_z}{E - p_z} \right) = \tanh^{-1} \left(\frac{p_z}{E} \right). \tag{3.4}$$

Under a boost in the z-direction to a reference frame with velocity β, the rapidity transforms as $y \to y - \tanh^{-1} \beta$, so that Δy is invariant under z-direction boosts. This property implies that the shape the rapidity distribution, $\mathrm{d}N/\mathrm{d}y$, and the opening angle between two objects, are also invariant against such boosts. When $p \gg m$ and $\theta \gg 1/\gamma$, the dependence of rapidity on mass and momentum can be neglected, and it can conveniently be expressed as a function of the polar angle θ. In this regime, the rapidity converges to the pseudo-rapidity coordinate, η, given by:

$$\eta = -\log \tan(\theta/2) \approx y. \tag{3.5}$$

In cases where the approximation is no longer valid, the pseudo-rapidity is still a meaningful coordinate, but without the benefit of invariance in $\Delta\eta$ against longitudinal boosts. In the pseudo-rapidity coordinate, the CMS detector extends from $|\eta| = 0$ ($\theta = 90°$) in the central region to $|\eta| = 5.2$ ($\theta = 0.6°$) in the forward region.

Kinematic quantities at the LHC are often projected onto the x-y transverse plane, where longitudinal boosts can be neglected. The transverse momentum, p_T, is given by: $p_\mathrm{T} = \sqrt{p_x^2 + p_y^2} = p \sin\theta = p/\cosh\eta$. Other transverse projections are denoted with a similar 'T' subscript.

3.2.2 Inner tracking system

The tracking system at CMS is the inner-most component of the detector [15, 16]. The layout of the tracker is shown in Fig. 3.5. It surrounds the IP with a cylindrical

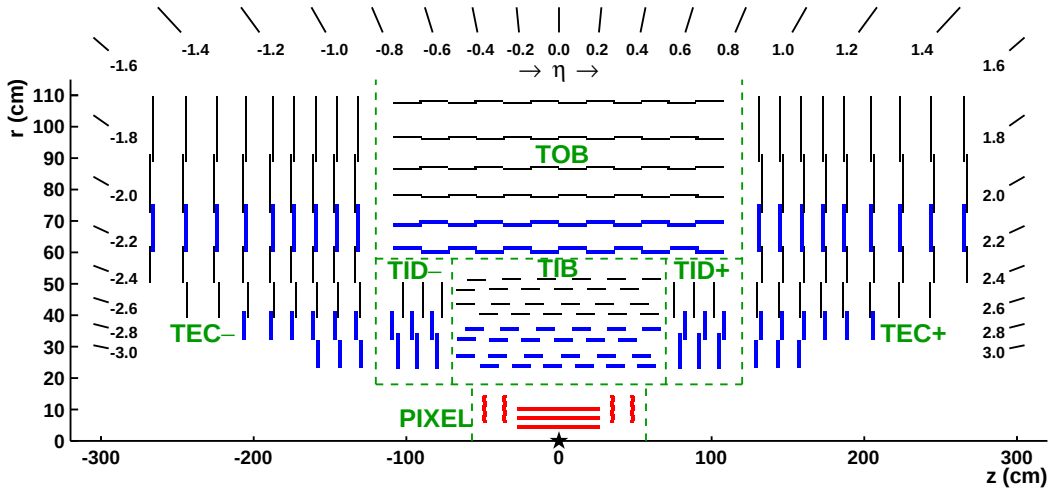

Figure 3.5: Schematic cross section of the CMS tracker in the r-z plane [14]. The pp interaction point, indicated by a star, corresponds to the center of the tracker. The pixel layer is shown in red. The strip tracker is indicated by the black and blue lines.

geometry beginning at just 4 cm from the beam line and extending outward to radius of 1.25 m. In the longitudinal direction, it has a length of 5.8 m, allowing for an angular coverage of $-2.5 < \eta < 2.5$ for charged particles with $p_{\mathrm{T}} > 1\,\mathrm{GeV}$. The tracker is contained inside a homogeneous 3.8 T magnetic field provided by the CMS solenoid.

The tracker is designed to provide a precise reconstruction of charged particle trajectories and interaction vertices, while also withstanding the extreme conditions near the IP. It must have sufficient speed and granularity to enable the reconstruction of tracks in a high particle flux and high pileup environment. At the same time, its material budget must be minimized in order to limit the rate of multiple scattering, bremsstrahlung, photon conversion, and nuclear interactions in the tracking volume.

The tracker design is optimized to strike a balance between the competing set of goals outlined above. A high-granularity silicon pixel detector is located in the tracker's inner-most section, with three barrel layers located between 4.4 cm and 10.2 cm from the beam line. With individual pixels only $100 \times 150\,\mu$m in size, a manageable occupancy of 10^{-4} per pixel per beam crossing is achieved. At larger radii, a lower-granularity silicon strip detector containing 10 barrel layers extends outwards to 1.1 m from the beamline. The silicon strip detector covers a larger surface area in a region with a lower particle flux. The silicon strips have lengths between 10-20 cm and pitches between $80 - 180\,\mu$m, depending on their location in the tracking volume. Hit occupancies are maintained at the level of several percent. Both pixel and strip detectors contain endcaps, which extend the angular coverage of the tracking system up to $|\eta| = 2.5$. In total, the CMS tracker design used in Run 1 of the LHC is composed of 1440 pixel modules and 15148 strip detector modules, providing about 200m^2 of active area. It is the largest silicon tracker ever built.

Pixel detector

The pixel detector is crucial for the reconstruction of secondary vertices due to b quark and τ lepton decays. Given B meson and τ lifetimes of order 10^{-13}-10^{-12} s, an impact parameter resolution of at least $100\,\mu$m is needed to perform the reconstruction. The impact parameter resolution is primarily determined by the position resolution of the detector, as well as the extrapolation uncertainty of a track from the detector back to the track's origin. The former is addressed by using small pixel dimensions, and by utilizing charge sharing, illustrated in Fig. 3.6, between neighboring pixels to further improve the position resolution of a track hit. The latter effect is due to multiple scattering in the beam pipe and detec-

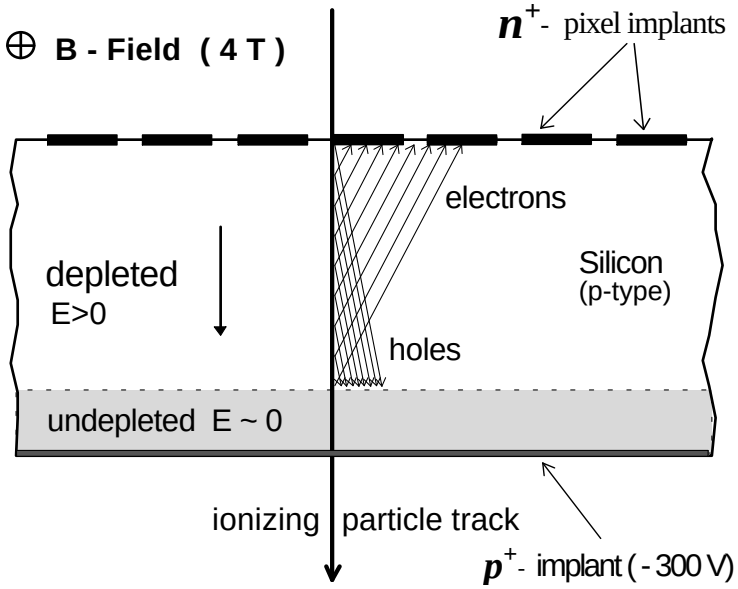

Figure 3.6: Illustration of charge sharing induced by $E \times B$ Lorentz drift in the pixel detector [15, 16].

tor material. Because an extrapolation is performed, the size of the uncertainty grows with distance from the beamline. Thus, the extrapolation uncertainty can be reduced by positioning the inner layer of the pixel detector as close as possible to the beamline, and by limiting the material budget of the detector to minimize multiple scattering.

The pixel detector provides a coverage of $-2.5 < \eta < 2.5$, with three barrel layers 56 cm in length at radii of 4 cm, 7 cm, and 11 cm from the beamline. Two disks are placed at either side of the barrel at $z = \pm 34.5$ cm and $z = \pm 46.5$ cm, each with an inner radius of 6 cm and outer radius of 15 cm. At distances near the beam pipe, every cm^2 of detector is expected to be impacted by 10 million particles per second at the nominal LHC luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$. The small pixel dimensions of $100\,\mu\mathrm{m}(r\phi) \times 150\,\mu\mathrm{m}(z)$ limit the occupancy for each pixel to 10^{-4} per pixel per beam crossing, and facilitate charge sharing between pixels to

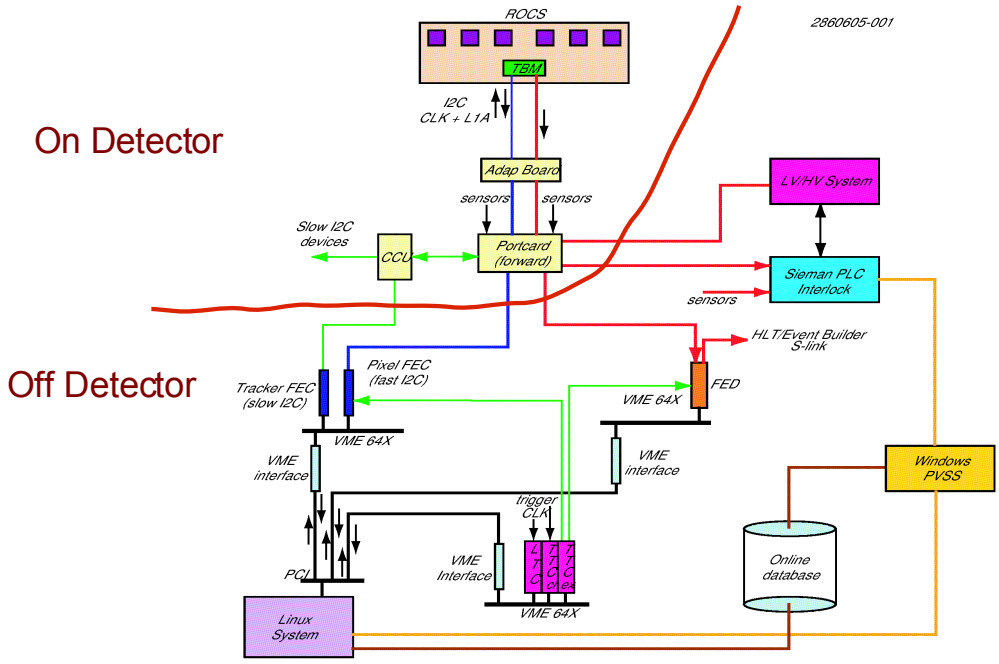

Figure 3.7: Schematic of the CMS pixel detector DAQ [17].

improve position resolution.

A schematic of the pixel data acquisition (DAQ) system is shown in Fig. 3.7. The detector contains 768 pixel modules in the barrel and 672 modules in the endcaps. Each module has up to 16 readout chips (ROCs) that are bump-bonded to the pixel sensors [50]. The ROC reads an array of 52×80 pixels in double-columns containing 160 individual pixels, and stores hit information for each pixel that exceeds a charge threshold. When a trigger arrives, the hit information is sent through the Token Bit Manager (TBM) to the Front End Driver (FED). Each TBM services 8 to 24 ROCs, distributing clock and trigger signals, and coordinating the readout among its ROCs. Each FED combines information from several TBMs, and sends it to the central DAQ. A portcard connected to each TBM transmits

information to the FED and receives information from the Front End Controller (FEC) through an optical link. The FEC sends triggers, clocks, and programming data to the TBMs and ROCs.

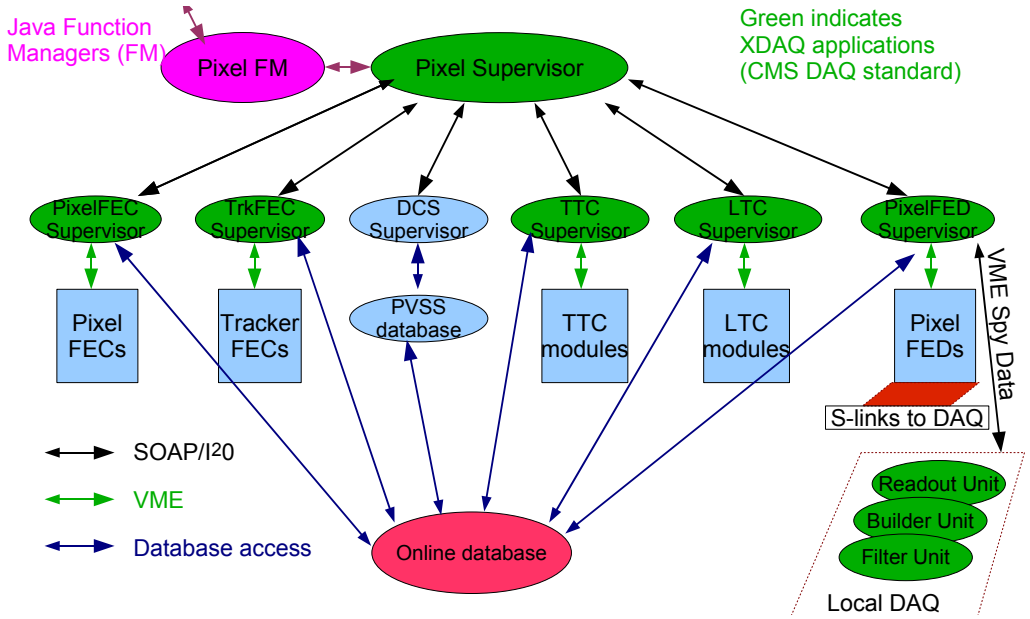

Figure 3.8: Schematic of the applications composing the POS [17].

The Pixel Online Software (POS) provides an interface to the pixel DAQ components. A schematic of the applications included in the POS is shown in Fig. 3.8. The software is based on the XDAQ platform developed at CERN for DAQ systems and widely used at CMS. The POS is composed of several supervisors that communicate directly with the respective pixel DAQ hardware components – for example, the PixelFEC Supervisor provides an interface to the FECs, and the PixelTTC Supervisor controls the TTC module for triggering and timing. The Pixel Supervisor is the top-level application that coordinates all pixel DAQ components. It runs the sequences necessary for the configuration and calibration of the pixel

detector.

Pixel phase-1 upgrade

The phase-1 upgrade targets the period starting with the second long shutdown at the LHC (LS2) and extends up to LS3 beginning in 2022 [51]. During this period, the LHC is expected to operate at luminosities up to 2-3 $\times 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ with 25 ns bunch spacing, which corresponds to an average of up to 50 pileup interactions per bunch crossing. The phase-1 detector must function in these conditions, and also withstand the cumulative radiation damage from an expected integrated luminosity of $300\,\mathrm{fb}^{-1}$ during this period.

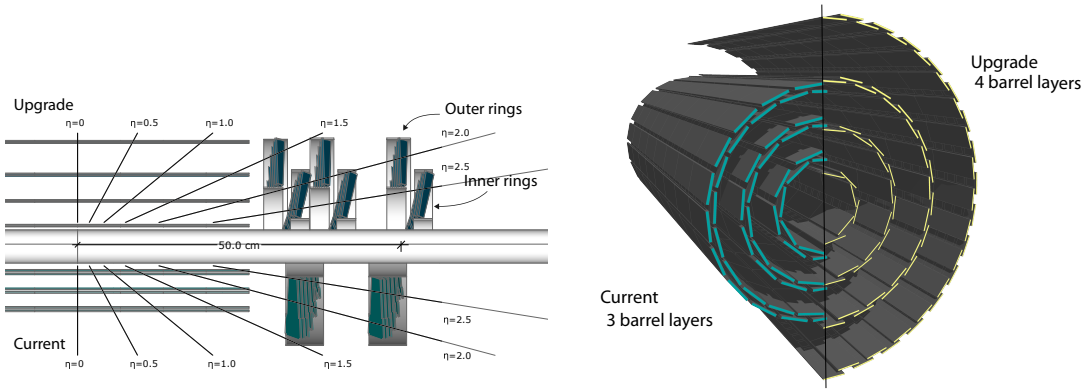

Figure 3.9: Layout of the pixel phase-1 detector compared to the original detector [18].

Improvements to the geometry of the detector include an additional barrel layer, extending coverage down to 3 cm from the beamline, and an additional forward disk. A comparison of the original and phase-1 detector is shown in Fig. 3.9. The total number of pixels will increase from 66 M to 124 M. At the same time, the material budget of the detector has been reduced. The phase-1 detector includes a new light-weight support structure, and a new cooling system that uses CO_2

instead of C_6F_{14}. In addition, part of the DAQ electronics are moved out of the tracking volume.

The phase-1 upgrade also includes updates to several components of the pixel DAQ, including a new ROC, TBM, and FED, along with new optical links. The updated ROC digitizes the encoding of the pixel address and pulse height data, doubles the readout buffer size to reduce dynamic data loss, and allows for lower noise thresholds (and higher radiation damage levels). The TBM is updated to accommodate the new digital signal from the ROCs. To accommodate a factor of two increase in the detector's power consumption, a larger voltage is delivered to the detector and then stepped-down using on-detector DC-DC converters. The phase-1 upgrade also changes the VME interface for the FED and FEC to a more efficient μTCA architecture.

A phase-1 pixel pilot blade was inserted into the current pixel detector during LS1 to help gain experience with the new detector. The pilot blade was inserted in the forward portion of the detector behind the existing two forward disks. The full phase-1 detector is expected to be installed during the 2016 year-end technical stop. Debugging and calibration exercises on the pilot blade are currently ongoing to ensure a seamless transition to the new detector.

3.2.3 Electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) is composed of lead tungstate ($PbWO_4$) crystals that release scintillation light when impacted by electromagnetic particles [19]. The ECAL is a homogeneous calorimeter in which the $PbWO_4$ crystals simultaneously produce particle showers and measure their energy output. It is

also hermetic, providing full coverage around the IP for the region with $|\eta| < 3$. The layout of ECAL is divided into a barrel section (EB) containing 61200 crystals, two endcaps (EE) with 7324 crystals each, as well as a preshower detector (ES) positioned in front of the endcaps. The EB covers the pseudo-rapidity range $|\eta| < 1.479$, and the EE cover $1.479 < |\eta| < 3.0$. A schematic of the detector is shown in Fig. 3.10. The crystal arrangement is designed to prevent cracks between crystals to be aligned with the IP – the EB crystals are mounted with a 3° slant with respect to the nominal IP, and the EE crystals are arranged in a 'Dee' shape with their faces pointing at a focus point located 1300 mm beyond the IP. Lead tungstate crystals have a high density of 8.28 g/cm^3, allowing for fast timing, fine granularity, and good radiation resistance. About 80% of the light produced in a crystal is emitted within the 25 ns window set by the LHC bunch crossing rate. The short radiation length of 0.89 cm and small Molière radius of 2.2 cm result in contained particle showers that allow for a fine granularity.

The ECAL preshower is located in front of EE with a coverage of $1.653 < |\eta| < 2.6$ to identify neutral pions impacting the detector. It is a sampling calorimeter consisting of two layers, where lead radiators initiate electromagnetic showers, and silicon strip sensors positioned behind each radiator to measure the shower's energy and transverse profile. The total thickness of the preshower is 20 cm.

Blue-green scintillation light in the lead tungstate crystals is produced with a wavelength in the range 420-430 nm, and then collected by avalanche photodiodes (APDs) in EB and vacuum photodiodes (VPDs) in EE. The choice of photodiodes is optimized for the magnetic field and radiation levels incident on the detector. The APDs have a better quantum efficiency, providing a higher incident photon to converted electron ratio, and the VPDs have a larger surface coverage on the

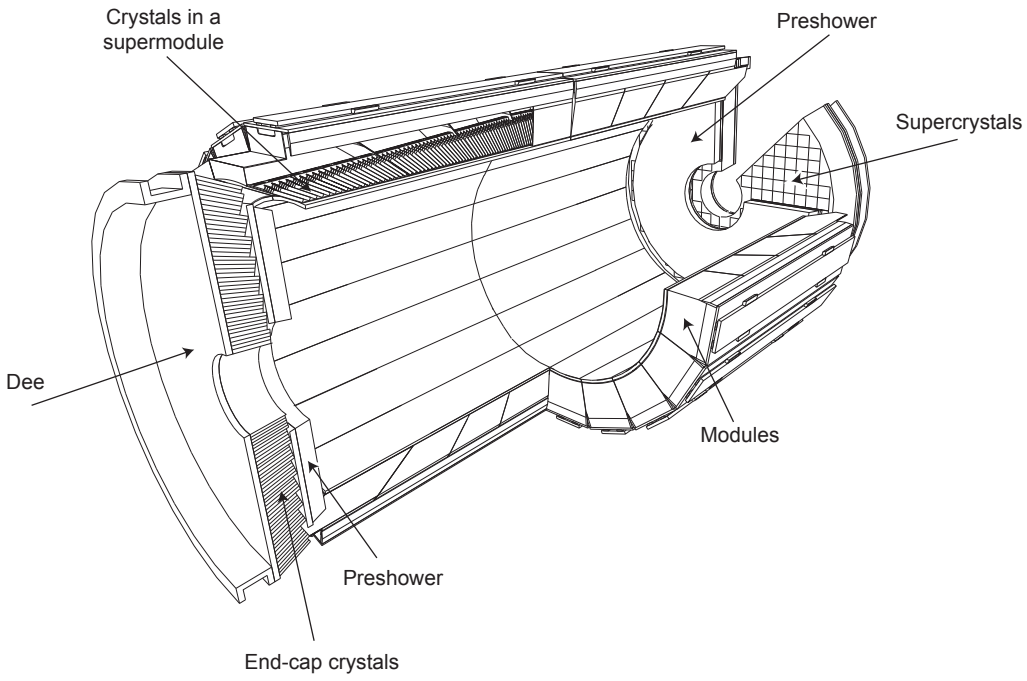

Figure 3.10: Schematic illustration of the CMS ECAL [19].

back face of each crystal. The crystals are polished on all faces with the exception of one unpolished lateral face that improves light uniformity in each crystal.

The ECAL is instrumental in reconstructing $H \to \gamma\gamma$ events and providing a good diphoton mass resolution. To achieve an energy resolution at the level of a few per mille, many small effects in the crystal transparency and light collection efficiency must be addressed. Exposure to ionizing radiation creates a wavelength-dependent loss of transparency over time due to the formation of color centers in the crystal volume. These effects are minimized by operating the detector at a temperature of 18° C, where an equilibrium is reached between the radiation damage and the annealing rate. For a constant dose rate, an equilibrium transparency can be maintained at this temperature. In practice, the transparency has a cyclic

behavior due to the intermittent nature of LHC collision runs and machine refills. The magnitude of these variations range between 1-2% in the barrel, and tens of percent in the high-η regions. For this reason, the crystal transparency must be monitored and corrected for in real time using injected laser light at blue (440 nm) and near infra-red (796 nm) wavelengths. The temperature of 18° is maintained with a water cooling system incorporated into the detector.

In addition to laser monitoring of the crystal transparency, ECAL employs several calibration schemes to mitigate global and channel-to-channel offsets in the energy scale. The latter is due to crystal-to-crystal variation of the scintillation light yield, and the yield collected by the photodiodes. Calibrations are conducted using cosmic ray muons, and also *in situ* using $\mathrm{W} \to \mathrm{e}\nu$, $\pi^0 \to \gamma\gamma$, and $\eta \to \gamma\gamma$ events.

After calibration, the resolution on energy E measured in ECAL can be parametrized by:

$$\left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C^2. \tag{3.6}$$

Here, S is a stochastic term describing event-to-event fluctuations in lateral shower containment, photostatistics, and energy deposited in ES; N is a noise term parametrizing non-uniformities in longitudinal light collection, intercalibration errors, and leakage of energy from the back of the crystals; and C is a constant term for electronics, digitization, and pile up noise. Typical values for these parameters are: $S = 2.8\%$, $N = 0.12$, and $C = 0.30\%$.

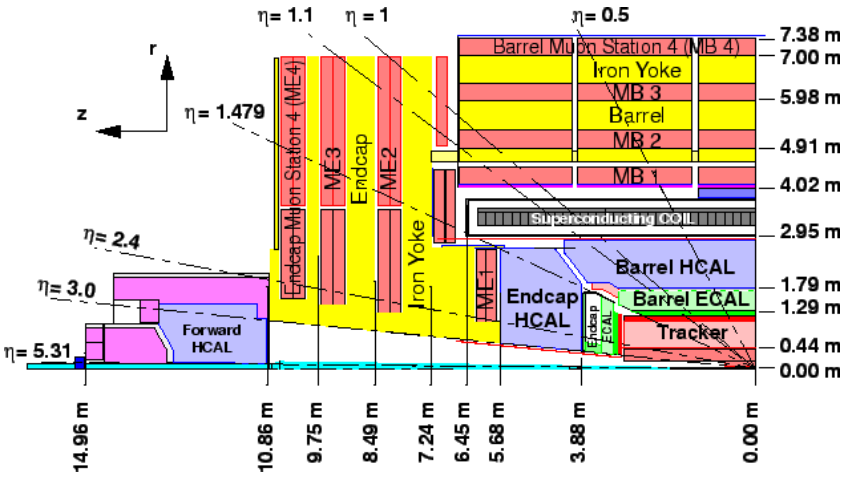

Figure 3.11: Schematic cross section of CMS with HCAL and muon sections visible [20].

3.2.4 Hadronic calorimeter

The hadronic calorimeter (HCAL) is a sampling calorimeter with alternating layers of absorber for initiating showers and scintillator for measuring their energy [21]. Radially, HCAL occupies the space between ECAL, with an outer edge at $R = 1.77$ m, and the solenoid beginning at $R = 2.95$ m. The calorimeter includes a barrel section (HB) covering $|\eta| < 1.3$, and an endcap (HE) covering $1.3 < |\eta| < 3$. A tail catcher (HO) covering $|\eta| < 1.3$ is also placed outside the solenoid coils to measure hadronic showers that are not contained in the calorimeter core. For hadronic activity in the forward region, with $3 < |\eta| < 5.2$, a forward calorimeter (HF) is attached to the ends of CMS 11.2 m from the IP. A schematic of the HCAL is shown in Fig. 3.11.

The HB consists of 36 azimuthal wedges, each containing absorber plates parallel to the beam axis alternating with plastic scintillator tiles. The absorber plates are constructed primarily from brass, with the innermost and outermost plates in each wedge made with stainless steel to improve structural strength. At 90° from

the beam axis, the absorber is 5.82 interaction lengths (λ_I) thick, with the addition of about 1.1 λ_I of material due to EB; at the forward edge of HB ($|\eta| = 1.3$), the thickness is 10.6 λ_I. The arrangement of plastic scintillator tiles in each wedge into 16 η-sectors provides a segmentation of $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$. The tiles are connected to wavelength-shifting fibers that extract the scintillation light and ultimately direct it to a hybrid photodiode (HPD).

The HE is constructed from brass absorber plates layered with plastic scintillator tiles. Wavelength-shifting fibers collect scintillation light and direct it to HPDs. The arrangement of tiles provides a granularity of $(\Delta\eta, \Delta\phi) = (0.087, 0.087)$ for $|\eta| < 1.6$ and $(\Delta\eta, \Delta\phi) = (0.17, 0.17)$ for $|\eta| \geq 1.6$. Together with the ECAL endcaps, the HE extends to a length of about $10\lambda_I$ from the IP.

The HO is positioned outside the solenoid coils in order to detect hadron showers that have penetrated both EB and HB in the region $|\eta| < 1.3$. The solenoid coil acts as an absorber with $1.4/\sin\theta$ radiation lengths, so that the total length of the calorimeter system is extended to a minimum of $11.8\lambda_I$, except in the barrel-endcap boundary region. The HO is segmented into 12 identical ϕ-sectors, with some gaps in coverage due to cracks, support beams, and cryogenics. The HO has direct physics impact in events with large hadronic activity. It is especially important for the measurement of $\vec{\not{E}}_{\mathrm{T}}$ (Section 4.6), where energy not captured by the EB and HB would otherwise broaden the $\vec{\not{E}}_{\mathrm{T}}$ resolution.

The very forward region of CMS with $3 < |\eta| < 5$ is covered by the HF calorimeter. In this region, the detector is required to withstand extreme conditions due to high particle multiplicities. The calorimeter includes a steel absorber composed of grooved plates with a thickness of 5 mm. Quartz fibers are inserted into the groves for the detection of charged shower particles through their emission of Cherenkov

light. These fibers are mostly sensitive to the electromagnetic component of showers. To distinguish between electromagnetic showers, generated by electrons and photos, and showers generated by hadrons, the fibers are contained in bundles of two different lengths. Long fibers run over the full depth of the absorber (165 cm), and short fibers begin at a depth of 22 cm from the front of the detector. In this configuration, electromagnetic showers deposit most of their energy in the first 22 cm of the long fibers, and hadronic showers produce nearly equal energy deposits in both fiber lengths. When both fiber lengths are read out simultaneously, the electromagnetic and hadronic components of a shower can be decoupled.

3.2.5 Muon system

The muon system is the outer-most part of CMS [52]. It has a cylindrical geometry with a barrel section covering $|\eta| < 1.2$, and two endcaps covering $0.9 < |\eta| < 2.4$. The muon barrel consists of four stations forming concentric cylinders around the beam axis. The middle two cylinders are embedded in the magnet's iron return yoke. A schematic of the muon barrel layout is shown in Fig. 3.12. The low rates and relatively low local magnetic field strength in the barrel region enable the use of drift tubes (DTs) as the active detector element. A schematic of a DT chamber embedded inside the iron yoke is shown in Fig. 3.13. Each DT is composed of two or three superlayers (SLs), where each SL contains four layers of rectangular drift cells staggered by a half cell. Each drift cell has an anode wire at +3600 V running along its center axis, with electrode strips at +1800 V along the wide faces and cathode strips at −1200 V along the thin faces. The wires in the outer SLs are parallel to the beam, providing a track measurement in the r-ϕ plane; the wires in the inner SLs are orthogonal to the beam, providing a measurement in

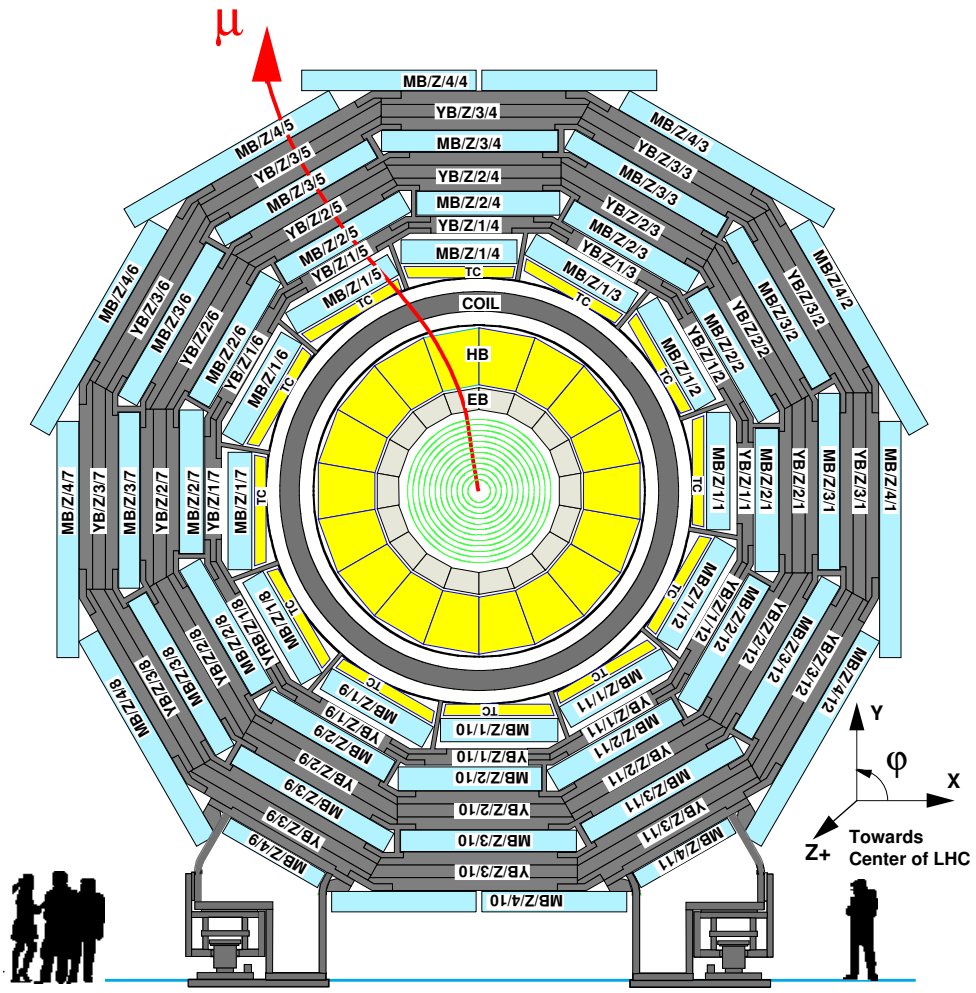

Figure 3.12: Cross section view of the CMS detector, showing the tracker in green, HCAL in yellow, magnet and return yoke in gray, and the muon barrel in light blue [21].

the z direction. The DTs include a gas mixture of about 85% Ar and 15% CO_2 by volume. About 10% of the gas is exchanged with fresh gas daily.

The muon endcap system is composed of Cathode Strip Chambers (CSCs), which are multiwire proportional chambers with six anode wire planes interleaved with seven cathode strips running in the perpendicular direction. Each chamber is filled with gas that is ionized by impacting muons. A schematic of the CSC

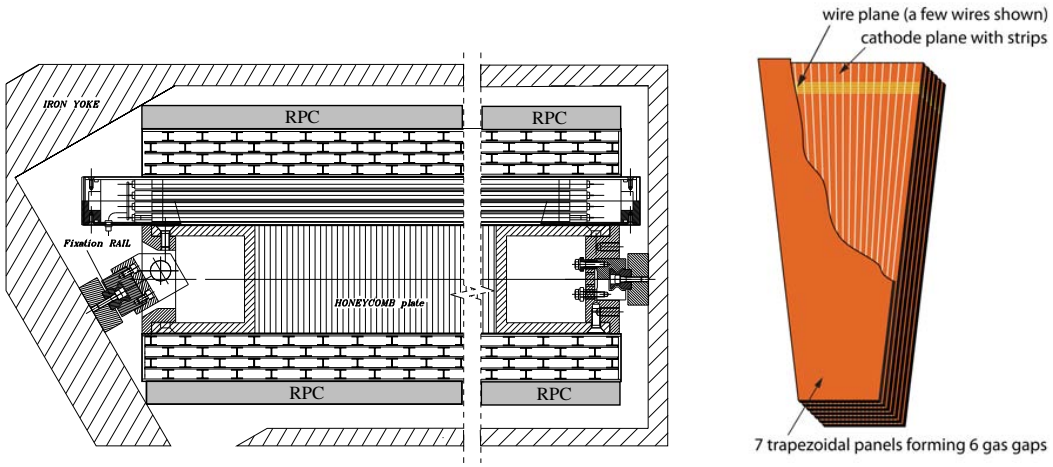

Figure 3.13: (Left) cross-section in the r-ϕ plane of a DT chamber embedded inside the magnet's iron return yoke. Two SLs are located on the top and bottom of the DT – the wires run along the beam direction (out of the page). One SL is oriented perpendicular to the beam direction. The honeycomb plate in the center is added for support. (Right) layout of a CSC with seven trapezoidal cathode panels and six gas gaps with planes of sensitive anode wires [21].

is shown in Fig. 3.13. The CSCs perform well in the high rates and large (and non-uniform) magnetic fields that are present in the endcap region. They do not require precise control of gas, temperature, or pressure for reliable operation.

To enhance the triggering capabilities of the muon system, Resistive Plate Chambers (RPCs) are inserted into the barrel and endcap regions. The RPCs are gaseous parallel-plate detectors that have a sufficient spatial resolution and fast timing comparable to that of scintillators. They allow for the tagging of an ionizing event in a time period that is much shorter than the 25ns LHC beam crossing rate. The RPCs are arranged in an approximate cylindrical geometry on 12 faces around the interaction point. Six layers of RPC chambers are embedded in the barrel, and 4 layers in the endcap cover a region up to $|\eta| = 2.1$.

3.2.6 Triggering and data acquisition

At the instantaneous luminosity of $10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, CMS encounters approximately 20 pileup interactions at intervals of every 25 ns. This corresponds to about 1 Mb of zero-suppressed data arriving at a rate of 40 MHz, orders of magnitude above the archival storage capacity available. The Trigger and Data Acquisition System (TriDAS) is responsible for filtering the data to a manageable rate of $\sim$ 100 Hz, and recording it for further analysis [22, 23].

The rejection rate of $\mathcal{O}(10^5)$ occurs in several stages in TriDAS. In the first stage, the Level-1 Trigger (L1T) reduces the 40 MHz bunch crossing rate to approximately 100 kHz. The L1T has approximately $3\,\mu$s to arrive at a decision, with the time limitation imposed by buffer sizes in the front-end electronics. Accounting for the finite latency of communication between the front-end electronics and the processing elements of the L1T, the computational time alloted is realistically no more than $\sim 1\,\mu$s. The data processed by L1T is therefore limited to course-granularity, low-resolution information from the local detector systems.

The L1T is composed of three major systems – the calorimeter, muon, and global triggers. A schematic of the L1T architecture is shown in Fig. 3.14. The calorimeter trigger combines information from the ECAL, HCAL, and HF detectors. The muon trigger is organized into systems corresponding to the RPC, CSC, and DT systems. The global trigger combines information from the entire event, and sends a trigger decision to the Trigger Timing and Control (TTC) system, which then signals the front-end and readout systems for each subdetector on the decision.

The second stage of filtering is carried out by the High-Level Trigger (HLT),

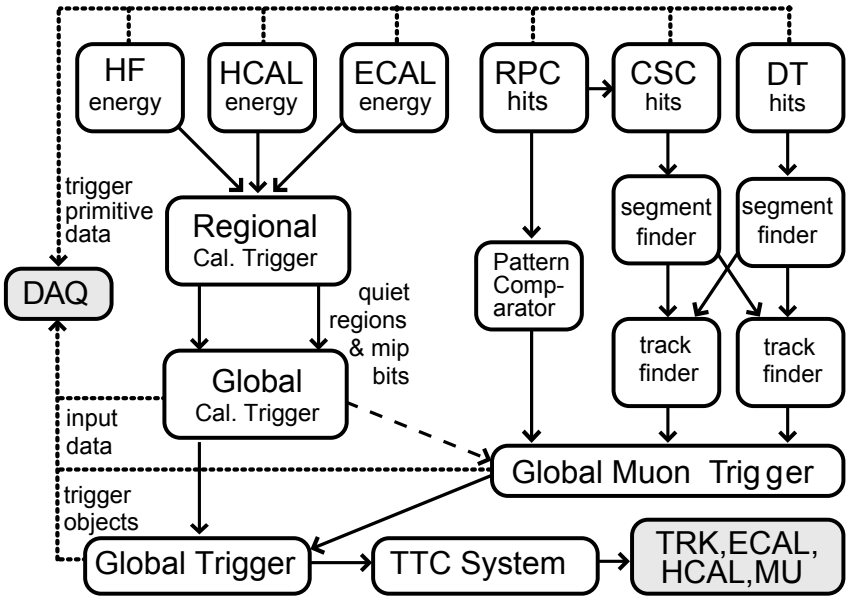

Figure 3.14: Schematic of the CMS L1T system [22].

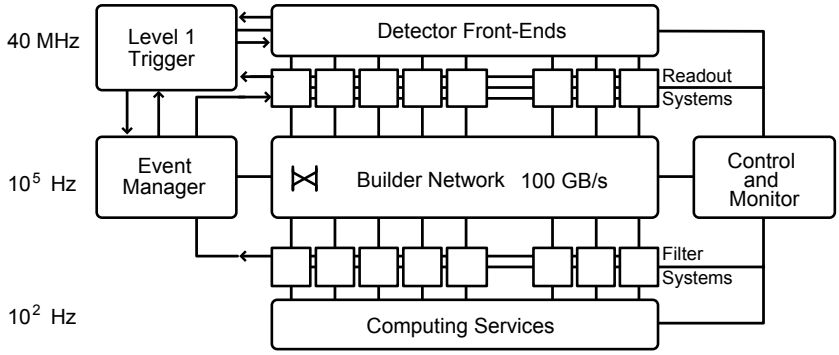

Figure 3.15: Architecture of the CMS DAQ system [22, 23].

with a rejection of $\mathcal{O}(10^3)$ to achieve the final rate of 100 Hz. In the $\sim$ 1 s allotted to HLT for a decision, this triggering stage employs more sophisticated physics algorithms that have access to reconstructed data from the entire detector. In addition to filtering events output by the L1T, the HLT categorizes events as candidates for various physics processes of interest.

A schematic overview of the DAQ architecture is shown in Fig. 3.15. Data from

the detector front-end electronics is stored in Readout Columns, each containing several FEDs and one Readout Unit that is responsible for buffering the event data. Data from the Readout Columns is routed through the Builder Network to one of $\sim$ 500 Builder Units, which assembles event fragments from the Readout Columns into a single event. The HLT algorithms are executed in the Filter Systems, which process the event data to make a triggering decision. Events passing the decision, as well as some events that fail, are passed on to the Computing Services for further analysis. In addition to these components, the Event Manager provides a centralized control of the event flow in the DAQ system, and the Control and Monitor system provides a user interface allowing for the configuration and monitoring of the DAQ.

CHAPTER 4

OBJECT RECONSTRUCTION

I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.

– William Thompson, 1st Baron Kelvin

Stable particles produced in proton-proton collisions inside CMS are identified and reconstructed by their pattern of energy deposits in the detector. Fig. 4 demonstrates how electrons, muons, photons, and hadrons are reconstructed in each subdetector system. Charged particles exhibit a helical trajectory in the solenoid's 3.8 T magnetic field, which is reconstructed by hits in the inner tracker. After passing through the tracker, electrons and photons deposit their energy in the ECAL. Muons typically penetrate all layers of the detector, and their momenta are reconstructed by matching tracks in the inner tracker to hits in the muon system. Charged and neutral hadrons deposit most of their energy in the HCAL. Undetected objects such as neutrinos or weakly-interacting BSM particles cannot be measured directly, but their presence can be inferred from the missing transverse momentum ($\vec{\not{E}}_{\mathrm{T}}$).

The particle-flow (PF) algorithm [53–55] is used at CMS to combine information from all of the subdetectors into fully-reconstructed objects such as electrons, muons, photons, and hadrons. The PF algorithm also enables jet reconstruction,

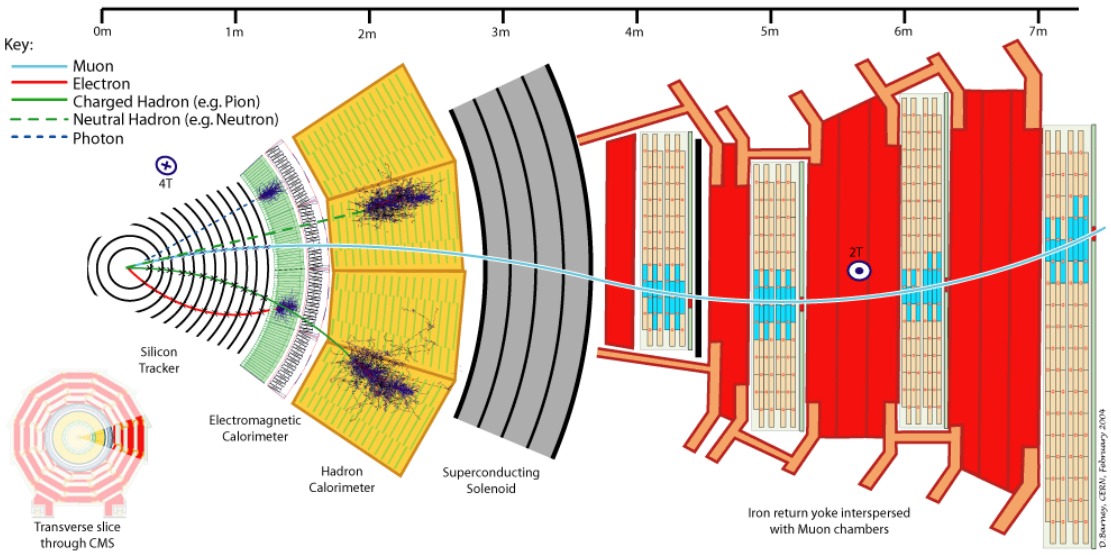

Figure 4.1: Schematic cross section of the CMS detector illustrating the interaction of stable particles with each of the subdetector elements [13].

where jet clustering is conducted using PF candidates instead of calorimeter energy deposits. The $\vec{\not{E}}_{\mathrm{T}}$ is reconstructed as a sum of all PF candidate momenta in the event. After reconstruction, corrections are derived for the PF objects using a combination of simulation and data control samples. These corrections may calibrate the object's energy scale, mitigate the effects of pileup, or address other sources of reconstruction bias.

4.1 Electrons

In the CMS detector, electrons produce hits in the inner tracker and deposit their remaining energy in the ECAL [56]. A fully-reconstructed electron contains a track matched to one or more ECAL energy clusters. The calorimeter clustering and track reconstruction algorithms employed for electron reconstruction are

designed to account for bremsstrahlung as the electron travels through the inner tracker layers. The fraction of the electron's energy contained in radiated photons ranges from 33% to 86%, depending on its η and the amount of tracker material it encounters. The photons are emitted tangentially to the electron's otherwise helical trajectory as it travels through the homogeneous 3.8 T magnetic field. This radiation causes sudden changes in the electron's curvature radius, and creates a spread along the ϕ-direction of its calorimeter energy deposit.

Several clustering algorithms are implemented at CMS for reconstructing electron energy deposits in the ECAL. In the PF method, clustering begins with a seed crystal exceeding a prescribed energy threshold, and aggregates all contiguous crystals with energies that are at least two standard deviations above the electronic noise. In this clustering method, a single ECAL crystal may be part of several PF clusters. This approach aims at reconstructing particle showers individually, as opposed to 'supercluster' methods that combine energy deposits from the electron and its radiated photons into a single cluster.

Charged particle tracks are typically reconstructed using a Kalman filter (KF) approach. For electrons, however, a dedicated approach is used to mitigate the effect of large radiative losses in the tracker material. In the first step, a seed consisting of two or three tracker hits is established. For PF electrons, the seed is found using tracks that have been reconstructed using the standard algorithm for generic charged particles, with appropriate matching to an ECAL cluster. Once a seed is established, track building is conducted using a combinatorial KF method, where layers of the tracker are added iteratively to obtain all of the hits corresponding to a track. Given a set of hits, a Gaussian Sum Filter (GSF) fit is performed to estimate the relevant track parameters. In this fit, the energy loss in each layer

of the tracker is estimated by a mixture of Gaussian distributions.

The PF electron is reconstructed by matching its GSF track to one or more PF clusters corresponding to the electron and any radiated photons emitted tangentially to its trajectory. The PF cluster corresponding to the electron is found by extrapolating the GSF track from its outer-most hit to the ECAL surface. To recover bremsstrahlung photons, straight lines tangent to the electron's trajectory are extrapolated to the ECAL at each layer of the tracker. Each matching ECAL cluster is added to the electron. In some cases, a bremsstrahlung photon is emitted and then undergoes a conversion. To recover these photons, a dedicated MVA algorithm is used to select displaced KF tracks and associate them with the electron's PF cluster.

4.2 Photons

The reconstruction of photons in CMS involves the clustering of ECAL energy deposits, as well as the association of electron tracks that may result from photon conversion in the tracker [57]. These ECAL clustering algorithms make no distinction between photons and electrons, which both have energy deposits of similar size and shape. For this reason, electrons with well-defined momenta, such as those produced in $Z \to e^+e^-$ decay, can conveniently be employed for studies of photon trigger, reconstruction, and identification efficiencies, as well as the photon energy scale and resolution. The clustering of ECAL energy deposits proceeds after the ECAL calibrations (and intercalibrations) have been performed to correct for differences in the energy response throughout various regions of the ECAL.

Conversions are addressed in photon energy reconstruction if the conversion

occurs before the last three layers of the tracker. A conversion can be identified by its characteristic electron track-pair topology with a displaced vertex in the tracking volume, where the two electron momenta are approximately collinear. The PF algorithm accounts for fully reconstructed conversions, and avoids the corresponding deposits from being identified as charged hadrons.

4.3 Muons

Muons are reconstructed by matching tracks from the inner tracker to hits in the muon system [58]. The 3.8 T solenoid field allows for a good muon momentum resolution of $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}} \sim 1\%$ at 100 GeV and $\sigma(p_{\mathrm{T}})/p_{\mathrm{T}} \sim 10\%$ at 1 TeV. Two techniques for muon reconstruction are employed at CMS. The Global Muon reconstruction technique begins with a standalone-muon track in the muon system, and matches it to a track in the inner tracker. The Tracker Muon reconstruction technique begins in the inner tracker, where all tracks satisfying a momentum threshold are considered as possible muon candidates. Each of these tracks is propagated outwards to the muon system, taking into account energy losses and Coulomb scattering in the detector material. If the track matches to at least one muon segment, a Tracker Muon is identified. The Tracker Muon reconstruction technique has a higher efficiency for low-momentum muons ($p \lesssim 5$ GeV), because of the looser requirement of only one muon segment in the muon system. Overall, about 99% of muons produced in pp collisions at CMS within the geometrical acceptance of the muon system are reconstructed as either Tracker or Global muons, often qualifying as both types. Global and Tracker muons sharing the same track in the inner tracker are merged into a single candidate.

4.4 Jets

Hadronic jets are typically clustered using the infrared and collinear safe anti-k_{T} algorithm [59] with a cone size of $\Delta R = 0.4$ or 0.5 as implemented in the FastJet package [60]. For PF jets, the clustering is conducted using PF candidates rather than calorimeter energy deposits. All PF objects are included with no restrictions based on particle type or energy threshold. In addition to PF electrons, muons, and photons discussed in Sections 4.1-4.3, charged and neutral hadrons reconstructed with the PF algorithm are also included in jet clustering. Charged hadrons are reconstructed from tracks in the inner tracker. Neutral hadrons are reconstructed from energy deposits in the HCAL that are separated from tracks. Neutral hadrons that overlap with charged hadrons are identified as an excess calorimeter energy deposit with respect to the charged hadron tracks. Hadron energies are obtained from the ECAL and HCAL, and for charged hadrons, input on the hadron momentum is obtained from its track. Jet identification criteria are applied in order to limit the rate of 'fake' jets due to calorimeter and/or readout electronics noise.

In MC simulation, particle-level jets are clustered using all stable particles within the ΔR cone, with the exception of neutrinos. These particle-level jets are used in the derivation of jet energy corrections (JEC), which calibrate the energy scale of PF jets to be consistent with the corresponding particle-level jets. The statistical spread in momentum between a PF jet and its particle-level counterpart is known as the jet energy resolution (JER).

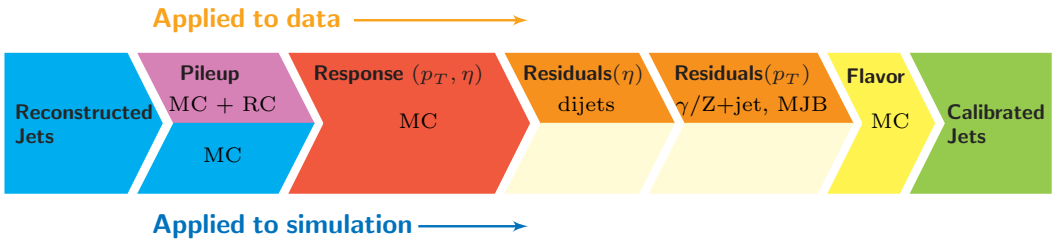

Figure 4.2: Consecutive stages of the JEC for jets in data and MC simulation [24].

4.4.1 Jet energy corrections

The JEC are derived using MC simulation, where jets reconstructed using PF objects are calibrated to the corresponding particle-level jets [24]. After the calibration in MC simulation, residual corrections are derived in order to account for differences between data and MC simulation. At CMS, the jet energy calibration uses a factorized approach, with several stages of corrections applied sequentially, as shown in Fig. 4.2 for jets in data and MC simulation. At each stage, the correction is applied as a constant multiplicative factor scaling the jet four-vector:

$$p_\mu^{\text{cor}} = \mathcal{C} \cdot p_\mu^{\text{raw}}, \tag{4.1}$$

where p_μ^{cor} is the corrected jet four-vector, p_μ^{raw} is the four-vector before corrections, and $\mathcal{C}$ is the correction factor. The value of $\mathcal{C}$ includes individual corrections for pileup (PU), response in MC simulation, and residual corrections accounting for differences between data and MC simulation. Each of these individual correction factors may depend on the jet p_T, η, flavor, and other jet-related quantities. The total correction is given as a product:

$$\mathcal{C} = C_{\text{offset}}(p_\text{T}^{\text{raw}}) \cdot C_{\text{MC}}(p'_\text{T}, \eta) \cdot C_{\text{rel}}(\eta) \cdot C_{\text{abs}}(p''_\text{T}), \tag{4.2}$$

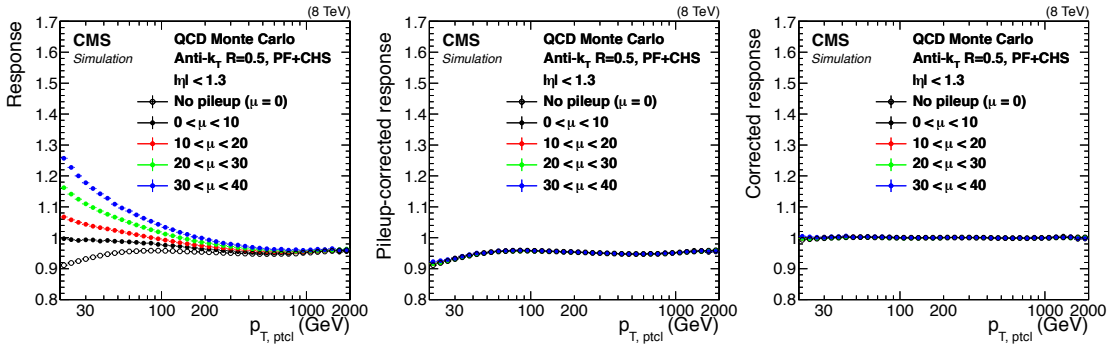

Figure 4.3: Ratio of reconstructed jet p_{T} to particle-level jet p_{T} in MC simulation. The ratio is shown for (left) uncorrected jets, (center) pileup-corrected jets, and (right) fully-corrected jets [24].

where p'_{T} and p''_{T} are the transverse momenta of the jet after the preceding corrections have been applied. This sequence of corrections is shown graphically in Fig. 4.2. The jet response for uncorrected jets, pileup-corrected jets, and fully-corrected jets is shown in Fig. 4.3. The determination of jet correction factors is summarized in more detail below.

Pileup offset correction

At CMS, PU occurs when multiple proton-proton collisions occur in a single event. In-time PU (IT PU) is a result of the high instantaneous luminosity provided by the LHC proton beams. Here, multiple proton-proton collisions in a single LHC beam crossing produce tracks and calorimeter deposits that do not correspond to the hard scatter of interest. The effects of IT PU can be partially mitigated by associating tracks to PU vertices, which are displaced from the signal vertex in the z-direction. Tracks, and matching calorimeter deposits, that are associated with PU vertices can be removed. However, neutral particles produced in the PU interactions cannot be identified directly, and the resulting calorimeter deposits

must be removed using other methods. In addition to IT PU, out-of-time PU (OOT PU) is due to the finite signal decay time in the calorimeters. Here, pp collisions from neighboring LHC bunch crossings may contribute to the calorimetric energy in the time window of a signal event. The contributions of OOT PU can be mitigated using signal processing techniques.

Charged particles produced in IT PU can be removed from jet clustering using the charged-hadron subtraction (CHS) technique. The leading primary vertex in the event is identified as the vertex corresponding to the largest sum of squares of track transverse momenta, $\sum |p_{\mathrm{T}}^{\mathrm{track}}|^2$. The remaining (subleading) vertices are classified as PU vertices. In the CHS method, charged hadron tracks are removed if they are associated with a PU vertex passing a set of quality requirements. The CHS technique is successful in removing approximately 50% of the IT PU within the tracker coverage. The remaining PU must be removed from jets using the PU offset corrections.

The hybrid jet area method is used to correct the jet energy scale for PU. The correction is parametrized with:

$$C_{\mathrm{hybrid}}(p_{\mathrm{T,raw}}, \eta, A_j, \rho) = 1 - \frac{\chi(\eta, \rho, \log(p_{\mathrm{T,raw}}))A_j}{p_{\mathrm{T,raw}}} \tag{4.3}$$

where χ encodes the dependence of the correction on η, the per-event PU p_{T} offset density ρ, and $\log(p_{\mathrm{T,raw}})$. The jet area, A_j, and ρ are calculated using the k_{T} clustering algorithm with distance parameter $D = 0.6$ and $|\eta| < 4.7$. They are determined by adding a large number of soft, nonphysical particles with infinitesimal momenta and random directions over the entire (η, ϕ) space. Jets with transverse momenta $p_{\mathrm{T}i}$ are clustered using these soft nonphysical particles. The offset energy in the event is defined as $\rho = \mathrm{median}(p_{\mathrm{T}i}/A_i)$. Here, the median is used to compute ρ instead of the mean in order to mitigate the quantity's sensitivity to

hard jets in the event.

The particle-level PU offset is determined in MC simulation by finding the average difference in p_{T} between matched jets in samples with and without PU overlay. To account for differences between data and MC, a scale factor is derived using a random cone (RC) offset method using zero-bias data and simulation. Here, the average p_{T} value of randomly-positioned cones in (η, ϕ) is taken to be the value of the RC offset. The RC offset is computed in both data and simulation, and the ratio between these quantities is used as a scale factor for the particle-level PU offset calculated in data.

Response correction

The jet response as a function of p_{T} and η is derived using MC simulation using a QCD multijet sample of 10 million events. The response is measured after the PU offset correction is applied. The particle-level response is defined as the ratio:

$$R_{\mathrm{ptcl}}(\langle p_{\mathrm{T}} \rangle, \eta) = \frac{\langle p_{\mathrm{T}} \rangle}{\langle p_{\mathrm{T,ptcl}} \rangle, \eta}[p_{\mathrm{T,ptcl}}, \eta], \tag{4.4}$$

where the ratio is derived in bins of particle-level p_{T} ($p_{\mathrm{T,ptcl}}$) and reconstructed η. The angle brackets $\langle\rangle$ indicate that averages of the p_{T} quantities are used within the relevant distribution bins. In the region $20\,\mathrm{GeV} < p_{\mathrm{T}} < 2\,\mathrm{TeV}$, reconstructed jets are corrected to within 0.5% with respect to the particle-level jet.

Residual corrections are derived to account for differences between jets reconstructed in data and MC simulation. Because particle-level information is not available in data, the residual response is determined using two different techniques that compare the p_{T} of a reconstructed jet to that of any recoiling activity.

In the p_{T} balance method, the jet p_{T} is compared to a reference object:

$$R_{\mathrm{jet},p_{\mathrm{T}}} = \frac{p_{\mathrm{T}\,\mathrm{jet}}}{p_{\mathrm{T}\,\mathrm{ref}}}, \tag{4.5}$$

where the reconstructed jet and reference jet are back-to-back in the lab frame. In the missing transverse momentum projection fraction (MPF) technique, the jet p_{T} is compared to the missing momentum in the event:

$$R_{\mathrm{jet,MPF}} = 1 + \frac{p_{\mathrm{T}}^{\mathrm{miss}} \cdot p_{\mathrm{T}\,\mathrm{ref}}}{(p_{\mathrm{T}\,\mathrm{ref}})^2}. \tag{4.6}$$

In both methods, part of the p_{T} imbalance may be due to the presence of additional jets in the event. To mitigate this effect, the corrections are determined as a function of additional jet activity in the event. The variable α is defined as a ratio: $\alpha = p_{\mathrm{T,3rd\ jet}}/p_{\mathrm{T\,ave}}$ for dijet events, and $\alpha = p_{\mathrm{T,2nd\ jet}}/p_{\mathrm{T},\gamma/\mathrm{Z}}$ for Z+jet and γ+jet events. The corrections are then extrapolated to $\alpha = 0$, where the relevant response effects can be isolated. A sample of QCD dijet events is used to determine the residual η-dependent response correction, and a sample of γ/Z+jet events is used to determine the residual p_{T}-dependent response correction.

JEC uncertainties

The uncertainties due to the determination of the JEC are a dominant source of systematic error in typical top quark mass measurements. The final JEC derived for the 8 TeV dataset have uncertainties below 1% for most central jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. They are summarized in Fig. 4.4.

The uncertainties due to the PU offset correction stem from the offset p_{T} dependence that is derived using MC simulation, and the scale factor correcting for an η-dependence in the data offset. Of these uncertainties, the former constitutes

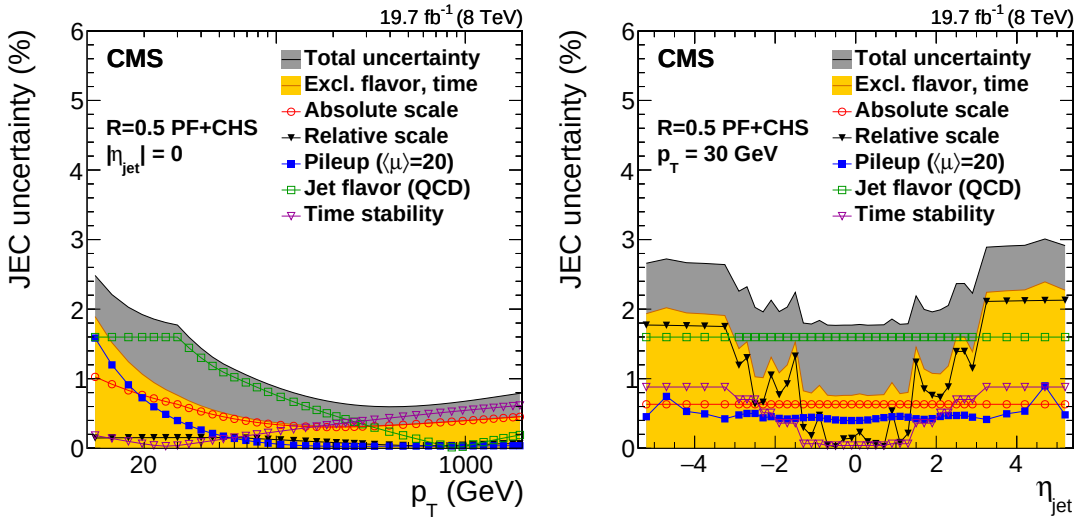

Figure 4.4: Uncertainties as a function of (left) jet p_{T} and (right) jet η for the JEC derived using the 8 TeV dataset. The relevant total uncertainty is shown in yellow, where corrections for flavor and time dependence are excluded. It is within 1% for most central jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ [24].

the dominant contribution across most of phase space. Other significant contributions to the JEC uncertainty stem from the residual corrections for the η- and p_{T}-dependent jet response. They include uncertainties in MC modeling, mostly stemming from initial and final state radiation, as well as statistical uncertainties in data samples and data-driven fits. The uncertainty sources are documented in detail in Ref. [24].

4.4.2 Jet energy resolution

The JER for central jets is typically 15-20% at 30 GeV, about 10% at 100 GeV, and about 5% at 1 TeV. It is defined as the width of the particle-level response

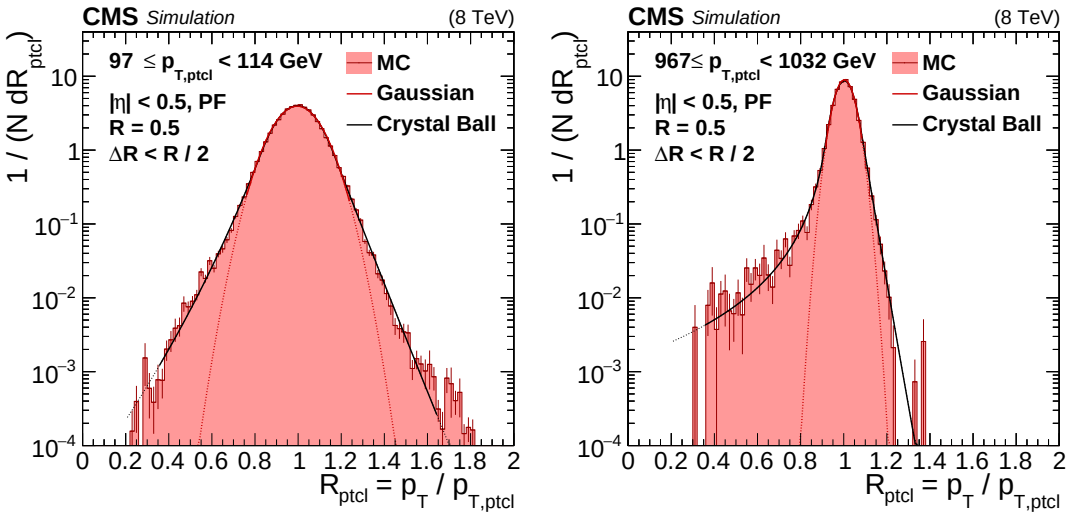

Figure 4.5: Distributions of the jet p_{T} resolution for central jets in two bins of p_{T}. The Gaussian core and power-law tails are well-modeled by a Crystal Ball function [24].

distribution,

$$R_{\mathrm{ptcl}} = p_{\mathrm{T,reco}}/p_{\mathrm{T,ptcl}}, \tag{4.7}$$

which is typically fit to a Gaussian distribution in its core region. Non-Gaussian effects become significant at approximately three standard deviations from the mean, and they can be parametrized using power law tails attached to the Gaussian core. Typical distributions of R_{ptcl} are shown in Fig. 4.5. The dependence of the JER on the jet p_{T} and the true number of pileup interactions, μ, is shown in Fig. 4.6. Understanding the JER is important for steeply falling spectra and resonant decays, where the smearing of jets can be a significant factor. The JER are also a central component in fake $\vec{\not{E}}_{\mathrm{T}}$, where mismeasured jets may cause a momentum imbalance faking the presence of undetected particles such as neutrinos. In many analyses, differences in the JER between data and MC simulation are a source of systematic uncertainty.

The particle-level JER is determined in MC simulation using QCD dijet events

[24]. Scale factors accounting for differences between data and simulation are also determined using QCD dijet events, with γ+jet events used for a complementary measurement. In these measurements, smearing can be caused by the JER in addition to other effects, such as the underlying event, out-of-cone showering, initial and final state radiation, and resolution effects corresponding to the reference object (a jet or photon). These effects are estimated and subtracted from the total resolution smearing to isolate the effects of the JER. The JER in MC simulation can be parametrized by the relation:

$$\frac{\sigma_{p_\mathrm{T}}}{p_\mathrm{T}} = \sqrt{\frac{\mathrm{sgn}(N)N^2}{p_\mathrm{T}^2} + \frac{S^2}{p_\mathrm{T}} + C}, \tag{4.8}$$

where N is a term incorporating the effects of noise and PU, S accounts for stochastic fluctuations that scale as $1/\sqrt{E}$, and C is a constant term. The noise term is found to have a large sensitivity to PU, contributing about 1 GeV of smearing (in quadrature) per additional PU vertex for cone-size $\Delta R = 0.5$ jets. The JER also carry a dependence on jet flavor, with gluon jets wider than quark jets, and heavy-quark jets undergoing larger smearing due to neutrinos produced in semi-leptonic decays. The data/MC scale factors determined for the 8 TeV dataset are shown in Fig. 4.7. The values are within 25% of unity in the central region, with an increase in the ratio for forward jets.

4.5 b jet identification

The bottom quark, or b quark, is in the same $SU(2)$ doublet with the top quark, making it a product of almost all top quark decays. It has a mass of ~ 4 GeV, and forms B mesons of mass ~ 5 GeV. The bottom quark decays through the weak interaction, and since it is lighter than the top quark, it must decay through a non-

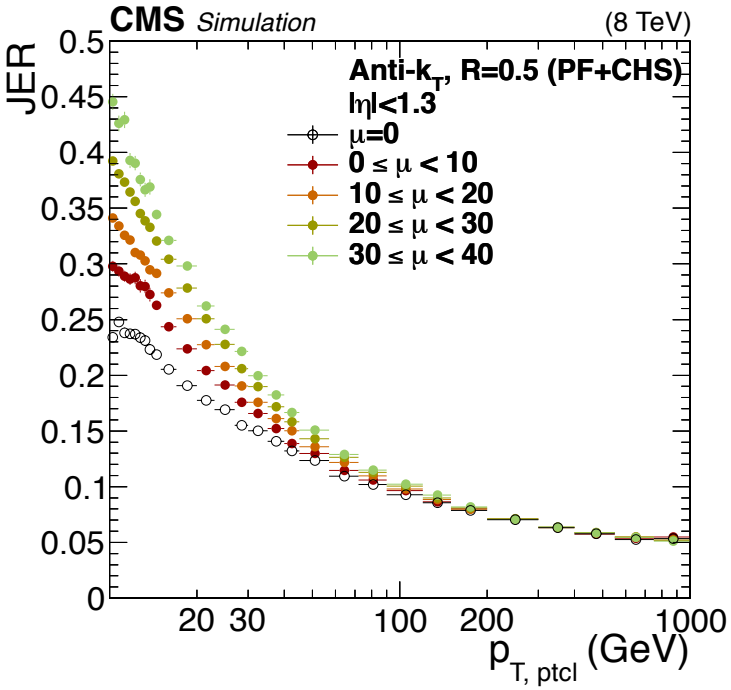

Figure 4.6: JER versus p_{T} for central jets in various bins of pileup μ [24].

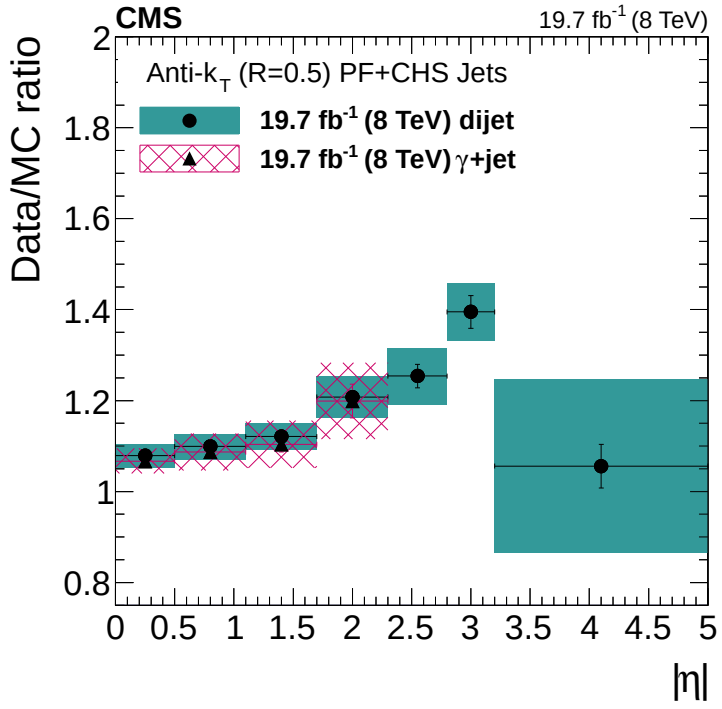

Figure 4.7: Data/MC scale factors and uncertainties measured using QCD dijet and γ+jet events using the 8 TeV dataset [24].

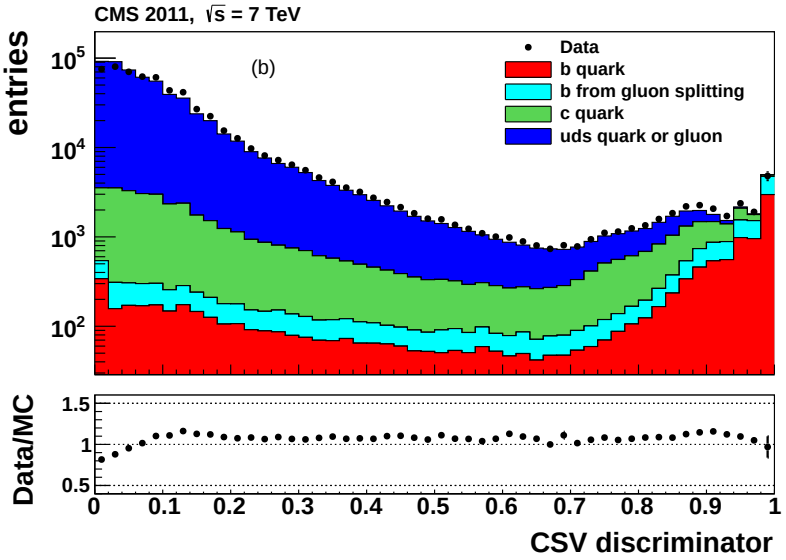

Figure 4.8: Distribution of CSV discriminator in QCD multijet events, with a breakdown of jet flavors in MC simulation [25].

diagonal element of the CKM matrix. For this reason, b decays have a relatively long lifetime of $\sim 10^{-12}$ s, corresponding to a decay length of $c\tau \sim 500\,\mu$m. This is large enough to create a discernible displaced vertex inside the tracking volume. (The charm quark has a slightly smaller lifetime of $\sim 10^{-13}$ s, making it more difficult, but also possible, to identify.) In addition to the characteristic displaced vertex in b decays, the large mass difference between the bottom quark and its decay products results in a hard spectrum of daughter particles. Finally, b decays typically produce a larger number of charged particles than other quarks, and have a semi-leptonic decay rate of $\sim 10\%$. Jets originating from the hadronization of bottom quarks are referred to as 'b jets'.

At CMS, b quarks are identified using b tagging algorithms that take advantage of the b decay properties outlined above. The central feature of these algorithms is the identification of tracks displaced from the primary vertex in the transverse plane. Such tracks can be identified by their impact parameter, or if there is a

sufficient number of displaced tracks, a fully-reconstructed secondary vertex corresponding to the b quark decay. The studies in this dissertation employ the Combined Secondary Vertex (CSV) algorithm developed at CMS [25]. This algorithm combines a set of variables with high discriminating power into a likelihood-based discriminator. The variables include properties of the secondary vertex – such as the transverse flight distance, mass, and number of tracks – as well as properties of the impact parameter and the jet itself. The distribution of the CSV discriminator for udsg and heavy-quark jets in QCD multijet events is shown in Fig. 4.8. The 'loose' operating point of the CSV b-tagging algorithm provides a $80-85\%$ b tagging efficiency, with a misidentification probability of about 10%.

4.6 Missing transverse energy

The PF-based missing transverse momentum is defined as a sum over all reconstructed PF objects:

$$\vec{\not{E}}_{\mathrm{T}} \equiv -\sum \vec{p}_{\mathrm{T}}, \tag{4.9}$$

which corresponds to the total transverse momentum of all unobserved particles, such as neutrinos or hypothetical weakly-interacting particles [26, 61]. The magnitude of $\vec{\not{E}}_{\mathrm{T}}$ is referred to as the missing transverse energy, or $\not{E}_{\mathrm{T}}$.

4.6.1 Missing transverse energy corrections

The PF $\not{E}_{\mathrm{T}}$ can be mismeasured due to minimum energy thresholds in the calorimeters, p_{T} thresholds and inefficiencies in the tracker, and nonlinearities in the calorimeter response. The first step in correcting the measured $\not{E}_{\mathrm{T}}$ is the propaga-

tion of JECs (Section 4.4.1) into the sum in Eq. 4.9, where a significant fraction of the PF candidates entering the sum may be contained inside a hadronic jet. The 'type-1' corrected $\vec{\not{E}}_{\mathrm{T}}$ is defined as:

$$\vec{\not{E}}_{\mathrm{T}}^{\,\mathrm{corr}} = \vec{\not{E}}_{\mathrm{T}} - \vec{\Delta}_{\mathrm{jets}} = \vec{\not{E}}_{\mathrm{T}} - \sum_{\mathrm{jets}} (\vec{p}_{\mathrm{T,jet}}^{\,\mathrm{corr}} - \vec{p}_{\mathrm{T,jet}}), \tag{4.10}$$

where the sum extends over all jets with an electromagnetic energy fraction below 0.9 and a corrected p_{T} exceeding 10 GeV. For each jet, $\vec{p}_{\mathrm{T,jet}}^{\,\mathrm{corr}}$ includes the p_{T} and η-dependent response corrections, but not the PU offset correction. The PU correction is excluded to preserve the isotropic nature of the PU in each event. Rather than implementing a jet-by-jet correction for PU, a more global event approach is implemented to correct the $\vec{\not{E}}_{\mathrm{T}}$.

The probability of producing neutrinos in inelastic proton-proton scattering is small, and PU is expected to contribute a negligible amount of genuine $\vec{\not{E}}_{\mathrm{T}}$. However, differences in the calorimeter response to charged and neutral particles cause the $\vec{\not{E}}_{\mathrm{T}}$ to point, on average, in the direction of the vectorial $\vec{p}_{\mathrm{T}}$ sum of the neutral particles. To mitigate this effect, a correction is derived as a function of the vectorial $\vec{p}_{\mathrm{T}}$ sum of all charged particles associated with PU vertices. The $\vec{\not{E}}_{\mathrm{T}}$ corrected for PU effects is given by:

$$\vec{\not{E}}_{\mathrm{T}}^{\,\mathrm{corr}} = \vec{\not{E}}_{\mathrm{T}} - \vec{\Delta}_{\mathrm{PU}} = \vec{\not{E}}_{\mathrm{T}} - \sum_{\mathrm{PU}} f(\vec{v})\vec{v}, \tag{4.11}$$

$$f(\vec{v}) = c_1(1.0 + \mathrm{erf}(-c_2|\vec{v}|^{c_3})), \tag{4.12}$$

where $\vec{v} = \sum_{\mathrm{charged}} \vec{p}_{\mathrm{T}}$ is the vectorial $\vec{p}_{\mathrm{T}}$ sum of charged PF candidates associated with a given PU vertex. Here, the factor $f(\vec{v})\vec{v}$ corresponds to the expected bias in $\vec{\not{E}}_{\mathrm{T}}$ stemming from each PU interaction. The values of the coefficients are measured to be $c_1 = -0.71$, $c_2 = 0.09$, and $c_3 = 0.62$, using simulated minimum bias events with exactly one generated proton-proton interaction.

After correcting the $\vec{\not{E}}_{\mathrm{T}}$ for jet response and PU effects, a small ϕ dependent asymmetry still exists due to a shift in the x and y components of $\vec{\not{E}}_{\mathrm{T}}$. The source of this asymmetry is a combination of imperfect detector alignment, reconstruction inefficiencies, a residual ϕ dependence in detector calibration, and a shift between the center of the detector and the beamline. Because PU is intrinsically isotropic, the magnitude of the asymmetry increases with increasing number of PU vertices. The correction for these effects is parametrized as a linear function of N_{vtx}, the number of reconstructed PU vertices:

$$\not{E}_{\mathrm{T}x}^{\mathrm{corr}} = \not{E}_{\mathrm{T}x} - \langle \not{E}_{\mathrm{T}x} \rangle = \not{E}_{\mathrm{T}x} - (c_{x_0} + c_{x_s} N_{\mathrm{vtx}}), \tag{4.13}$$

$$\not{E}_{\mathrm{T}y}^{\mathrm{corr}} = \not{E}_{\mathrm{T}y} - \langle \not{E}_{\mathrm{T}y} \rangle = \not{E}_{\mathrm{T}y} - (c_{y_0} + c_{y_s} N_{\mathrm{vtx}}), \tag{4.14}$$

where the coefficients c_{x_0}, c_{x_s}, c_{y_0}, and c_{y_s} are determined using $\mathrm{Z} \to \mu^+\mu^-$ events.

4.6.2 Large $\not{E}_{\mathrm{T}}$ due to misreconstruction

In addition to effects that cause bias in the reconstructed $\vec{\not{E}}_{\mathrm{T}}$, spurious detector signals may also be responsible for large $\not{E}_{\mathrm{T}}$ due to misreconstruction. These events are suppressed by four orders of magnitude from the core of the $\not{E}_{\mathrm{T}}$ spectrum, but cause large tails that must be addressed. Events with anomalous $\not{E}_{\mathrm{T}}$ are identified using detector-level information. Common causes include particles striking sensors in the ECAL, beam-halo interactions with the ECAL, dead detector cells, calorimeter and readout electronics noise, particle interactions with light guides and photomultiplier tubes, misfires of the HCAL laser calibration system, and incorrect track reconstruction. Addressing these effects mitigates the large $\not{E}_{\mathrm{T}}$ tails, and establishes good agreement between data and MC $\not{E}_{\mathrm{T}}$ spectra. The distribution of PF $\not{E}_{\mathrm{T}}$ in QCD dijet events with and without the cleaning algorithms is

shown in Fig. 4.9.

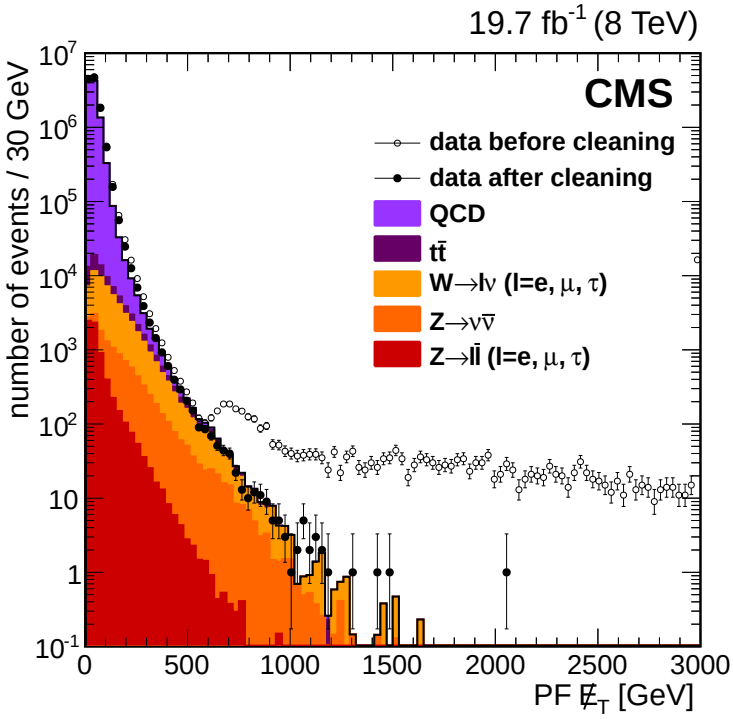

Figure 4.9: Distribution of PF $\not{E}_{\mathrm{T}}$ in QCD dijet events. The data distributions are shown before and after cleaning algorithms are implemented [26].

4.6.3 Missing transverse energy scale and resolution

The performance of PF $\vec{\not{E}}_{\mathrm{T}}$ reconstruction at CMS is assessed using Z +jet and γ+jet events, which are expected to be dominated by 'fake' $\vec{\not{E}}_{\mathrm{T}}$ due to object misreconstruction and resolution smearing. The magnitude of fake $\not{E}_{\mathrm{T}}$ is typically set by the resolutions of jets and unclustered hadronic activity. The values of typical JERs are shown in Fig. 4.6 – they are $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} \sim$ 15-25% at 25 GeV, and $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} \sim$ 10-15% at 50 GeV. The unclustered energy resolution is estimated in Chapter 5 to be $\sim 0.5\sqrt{\sum |\vec{p}_{\mathrm{T}}|}$, where the scalar p_{T} sum over unclustered PF candidates is of order 500 GeV. The momentum resolution of leptons and photons

is small relative to the hadronic components of the event. Typical resolutions are $\sigma_{p_T}/p_T \sim$ 1-4% for leptons reconstructed from Z-boson decay, and $\sigma_{p_T}/p_T \sim$ 1-3% for direct photons. The distribution of $\not{E}_T$ in $Z \to \mu^+\mu^-$ events is shown in Fig. 4.10. It is observed to peak at the characteristic $\not{E}_T$ resolution of $\sim 15\,\text{GeV}$.

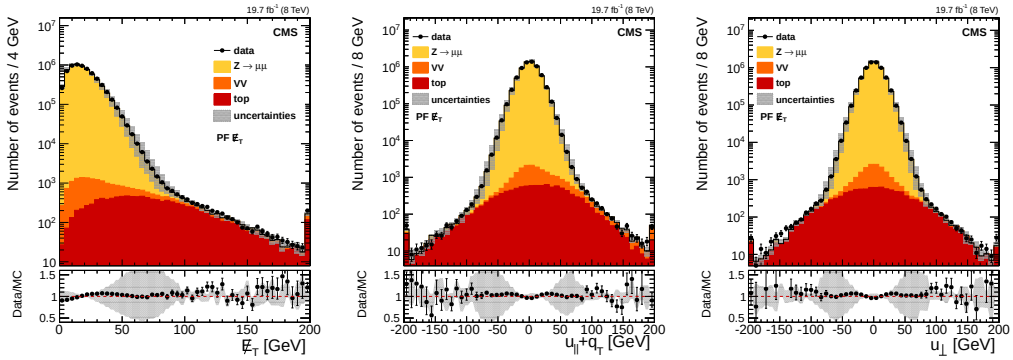

Figure 4.10: Distributions of (left) fully-corrected PF $\not{E}_T$ (center) $u_\parallel + q_T$, and (right) $u_\perp$ in $Z \to \mu^+\mu^-$ events. The data/simulation ratio is shown in the bottom panels, with systematic uncertainties on the simulation shown in grey [26].

To evaluate the scale and resolution of $\not{E}_T$, the well-measured momentum of a direct photon or $Z \to \ell^+\ell^-$ decay is compared to the recoiling hadronic activity in the event. The boson momentum in the transverse plane is denoted by $\vec{q}_T$. The hadronic recoil, defined as the vectorial sum of all reconstructed PF objects except the boson or its decay products, is denoted $\vec{u}_T$. By construction, we have that $\vec{q}_T + \vec{u}_T + \vec{\not{E}}_T = 0$.

The hadronic recoil in each event can be projected onto the boson momentum axis, along the direction of $\vec{q}_T$. This projection of $\vec{u}_T$ yields a parallel component, $u_\parallel$ and perpendicular component, $u_\perp$. These quantities are schematically illustrated in Fig. 4.11. In this formulation, the $\not{E}_T$ response is given by $-\langle u_\parallel \rangle / q_T$, where the negative sign accounts for the fact that the $\vec{u}_T$ is typically recoiling against the $\vec{q}_T$. The behavior of the response as a function of q_T is an important metric

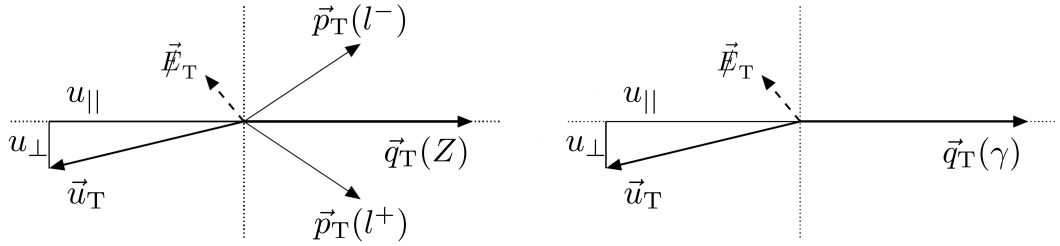

Figure 4.11: Schematic illustration of the (left) $\mathrm{Z} \to \ell^+\ell^-$ and (right) direct-photon kinematics in the transverse plane [26]. The $\vec{q}_{\mathrm{T}}$ vector corresponds to the boson's transverse momentum. The $\vec{u}_{\mathrm{T}}$ vector corresponds to the hadronic recoil, with parallel component $u_{\parallel}$ and perpendicular component $u_\perp$.

of $\vec{\not{E}}_{\mathrm{T}}$ performance, providing input on the $\vec{\not{E}}_{\mathrm{T}}$ reconstruction over a range of recoil momentum scales. The measured PF $\vec{\not{E}}_{\mathrm{T}}$ response is shown in Fig. 4.12. It is observed to be near unity at momentum scales above 40 GeV, with minor differences between physics samples due to differences in the quark and gluon jet composition of the samples. Below 40 GeV, the contribution of unclustered hadronic activity to the $\not{E}_{\mathrm{T}}$ becomes significant, since the unclustered activity is more-or-less isotropic and provides no recoil axis. Because the unclustered energy is not corrected for biases in energy scale, the response in this regime is underestimated.

The measured resolutions $\sigma(u_{\parallel})$ and $\sigma(u_\perp)$ are shown as a function of q_{T} in Fig. 4.13. The resolutions are observed to increase with increasing q_{T}, with the $\sigma(u_{\parallel})$ exhibiting an approximately-linear dependence and the $\sigma(u_\perp)$ curve increasing less rapidly. Studies of the $\vec{\not{E}}_{\mathrm{T}}$ resolution indicate an approximately-linear increase in the x and y components of $\vec{\not{E}}_{\mathrm{T}}$ as a function of $\sqrt{\sum |\vec{p}_{\mathrm{T}}|}$, the scalar p_{T} sum over all PF candidates in the event. The $\not{E}_{\mathrm{T}}$ resolution also carries a dependence on the number of PU interactions, with each additional interaction degrading the resolution by about 3.3-3.6 GeV, in quadrature [26]. In Chapter 5,

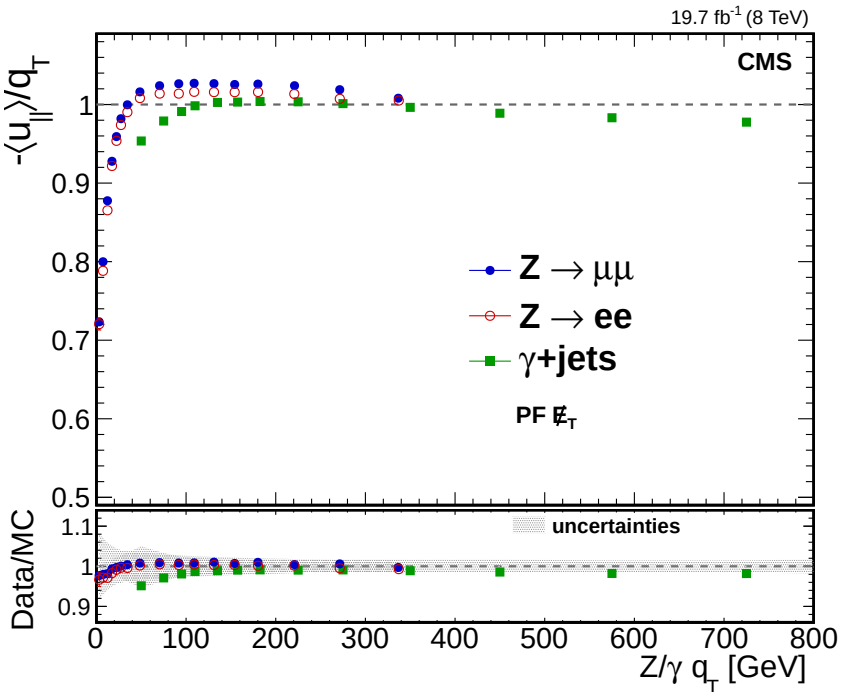

Figure 4.12: Response of PF $\vec{\not{E}}_{\mathrm{T}}$ in Z +jets and γ+jets events [26].

the $\vec{\not{E}}_{\mathrm{T}}$ significance algorithm is employed to estimate the $\vec{\not{E}}_{\mathrm{T}}$ resolution from the configuration of jets and unclustered hadronic activity in each event.

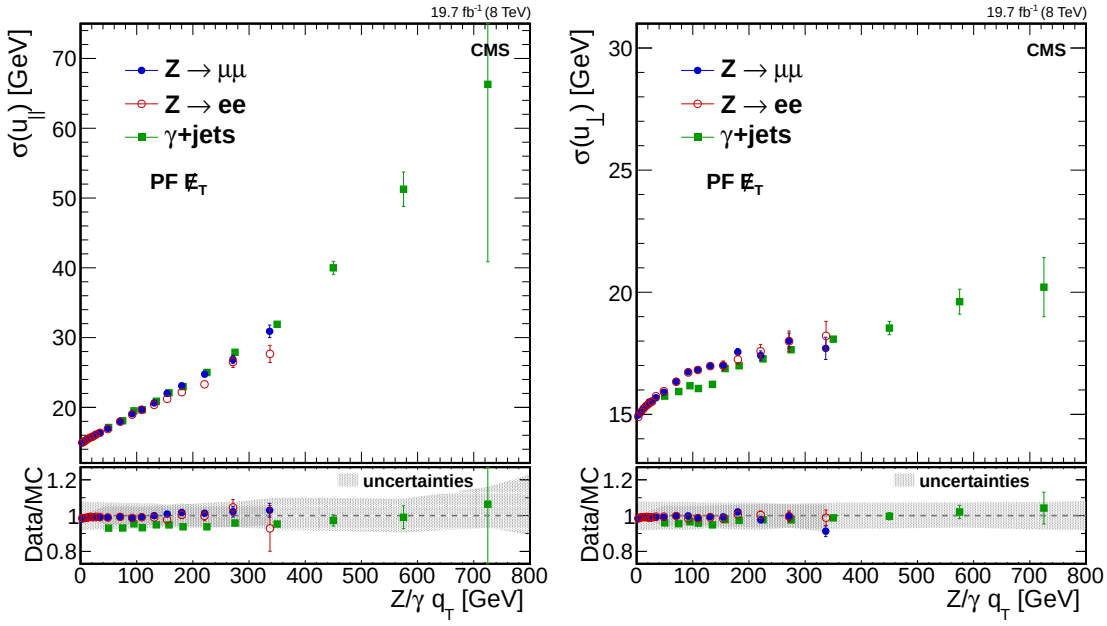

Figure 4.13: Resolutions of (left) $u_{\parallel}$ and (right) $u_{\perp}$ as a function of q_{T} in events containing a Z-boson decay or direct photon [26].

CHAPTER 5

$\vec{\not{E}}_T$ SIGNIFICANCE

Nothing is, but what is not.

– William Shakespeare, *Macbeth*

The ability to distinguish between events with spurious $\vec{\not{E}}_{\mathrm{T}}$ and those with genuine $\vec{\not{E}}_{\mathrm{T}}$ is important for analyses using missing transverse energy variables. Spurious $\vec{\not{E}}_{\mathrm{T}}$ may arise from object misreconstruction, finite detector resolution, or detector noise. To help identify such events, we have developed a missing transverse energy significance variable, which we will denote by '$\vec{\not{E}}_{\mathrm{T}}$ significance', or simply $\mathcal{S}$. On an event-by-event basis, $\mathcal{S}$ evaluates the p-value that the observed $\vec{\not{E}}_{\mathrm{T}}$ is inconsistent with a null hypothesis, $\vec{\not{E}}_{\mathrm{T}} = 0$, given the full event composition and resolution functions for each object in the event. A high value of $\mathcal{S}$ is an indication that the $\vec{\not{E}}_{\mathrm{T}}$ observed in the event is not well explained by resolution smearing alone, suggesting that the event may contain unseen objects such as neutrinos or more exotic weakly interacting particles. Characterstic events with spurious $\vec{\not{E}}_{\mathrm{T}}$ and genuine $\vec{\not{E}}_{\mathrm{T}}$ are illustrated in Fig. 5. The version of $\mathcal{S}$ described here is documented in Refs. [26, 62] and builds on previous algorithms from Refs. [61, 63].

5.1 Definition of $\mathcal{S}$

The significance is defined as the log-likelihood ratio,

$$\mathcal{S} \equiv 2 \ln \left(\frac{\mathcal{L}(\vec{\varepsilon} = \sum \vec{\varepsilon}_i)}{\mathcal{L}(\vec{\varepsilon} = 0)} \right). \tag{5.1}$$

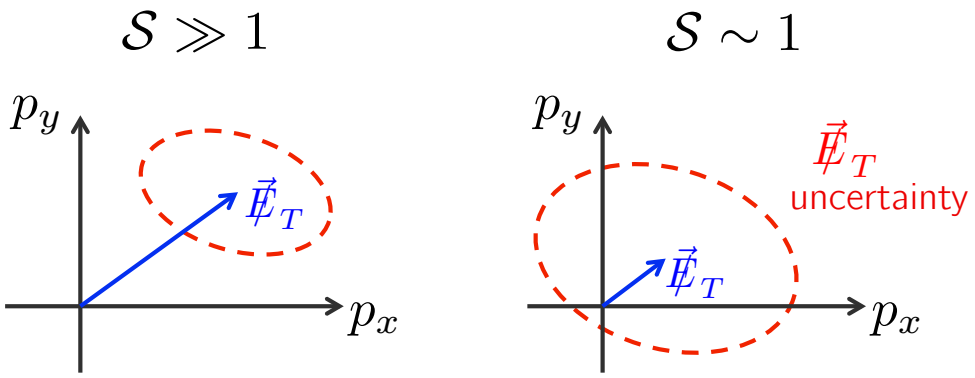

Figure 5.1: Illustration of a characteristic event with (left) nonzero true $\vec{\not{E}}_{\mathrm{T}}$ and (right) zero true $\vec{\not{E}}_{\mathrm{T}}$, where the observed $\vec{\not{E}}_{\mathrm{T}}$ is consistent with a null hypothesis. The observed $\vec{\not{E}}_{\mathrm{T}}$ is indicated by the blue arrow. The red dashed ellipse represents the contour $-2\Delta\mathcal{L}(\vec{\varepsilon}) = 1$, for $\mathcal{L}(\vec{\varepsilon})$ defined in Eq. (5.2).

The numerator expresses the likelihood of the hypothesis under test, that the true value ($\vec{\varepsilon}$) of the missing transverse energy is equal to the observed value ($\sum \vec{\varepsilon}_i$) , while the denominator expresses the likelihood of the *null hypothesis*, that the true missing transverse energy is actually zero. Under the null hypothesis, observation of any non-zero missing transverse energy is attributed to resolution smearing.

The formulation in Eq. (5.1) is completely general and accommodates any probability distribution functions for the object resolutions; throughout the bulk of this discussion however, we assume Gaussian resolutions for measured quantities. This assumption accurately describes the dominant behavior of energy and momentum measurements in CMS and greatly simplifies the computation of $\mathcal{S}$ as the convolution integrals underlying the likelihood functions can be done analytically. In the

Gaussian model, the likelihood $\mathcal{L}(\vec{\varepsilon})$ can be written:

$$\mathcal{L}(\vec{\varepsilon}) = \frac{1}{2\pi|\mathbf{V}|^{1/2}} \exp\left(-\frac{1}{2}(\vec{\varepsilon} - \sum_i \vec{\varepsilon_i})^\dagger \mathbf{V}^{-1}(\vec{\varepsilon} - \sum_i \vec{\varepsilon_i})\right), \tag{5.2}$$

in which $\mathbf{V}$ is the 2×2 covariance matrix of the total missing transverse energy computed by propagating the uncertainties of all objects in the event or in a defined subset of the event.

Evaluating the likelihood ratio in the Gaussian case, we obtain a simple closed-form solution,

$$\mathcal{S} = \Big(\sum \vec{\varepsilon_i}\Big)^\dagger \mathbf{V}^{-1} \Big(\sum \vec{\varepsilon_i}\Big). \tag{5.3}$$

Rotating into a coordinate system where the x-axis is aligned with the $\vec{\not{E}}_T$, $\mathcal{S}$ can be expressed simply as,

$$S = E_T^2/\sigma_{E_T}^2(1-\rho^2), \tag{5.4}$$

where $\sigma_{E_\mathrm{T}}^2$ is the diagonal variance of the measured $\not{E}_T$, and ρ is the correlation coefficient between the variances parallel to and perpendicular to the $\vec{\not{E}}_T$. The significance written in this way is reminiscent of the naïve version, $\Sigma = \not{E}_T/\sqrt{\sum E_\mathrm{T}}$, used in past CMS analyses. The strength of $\mathcal{S}$ lies in its comprehensive treatment of the denominator, where the configuration and resolutions of all reconstructed objects contribute individually to the estimated variance of the $\not{E}_T$ in each event.

A particularly useful feature of the Gaussian approximation is that the $\mathcal{S}$, as defined by Eq. (5.3), is a χ^2 variable with two degrees of freedom (one degree of freedom for each component of $\vec{\not{E}}_\mathrm{T}$). For clarity, we note that the term "significance" is often used to denote a linear quantity of the form x/σ_x while here it is defined as the quadratic form x^2/σ_x^2.

Despite the convenience of Eq. (5.3), a full treatment of $\vec{\not{E}}_\mathrm{T}$ significance must also include non-Gaussian resolutions as these are known to occur at the percent

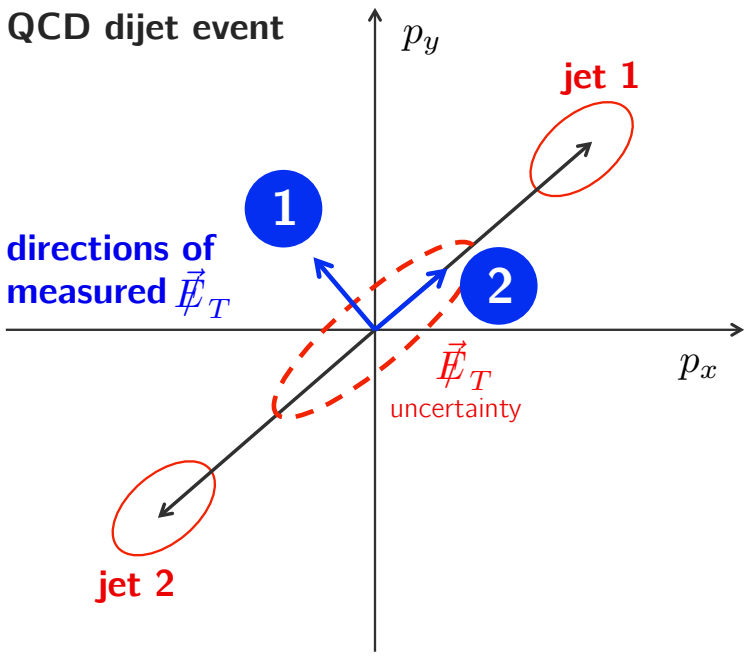

Figure 5.2: Illustration of the impact of event topology on the value of $\mathcal{S}$. In a hypothetical QCD dijet event, we compare the value of $\mathcal{S}$ in Case 1, where the $\vec{\not{E}}_{\mathrm{T}}$ is measured along the dijet axis, and Case 2, where the $\vec{\not{E}}_{\mathrm{T}}$ is perpendicular to the dijet axis. The $\not{E}_T$ is equal in both cases. The value of $\mathcal{S}$ is large in Case 1, and small in Case 2.

level in jet measurements. In Section 5.6 of this chapter we therefore extend the treatment of $\mathcal{S}$ to handle such cases.

5.2 Datasets and event selection

The core of the studies outlined in this chapter use events from a dataset recorded during 2012, corresponding to an intergrated luminosity of 19.7 fb^{-1} and

$\sqrt{s} = 8\,\text{TeV}$. The PF algorithm is used to reconstruct and identify each individual particle in an event. Electrons are identified with a simple cut-based approach. An isolation cut of $I_{\text{rel}} < 0.2$ is imposed with a cone size of $\Delta R = 0.4$. Muons are required to pass an isolation cut of $I_{\text{rel}} < 0.12$ with a cone size of $\Delta R = 0.4$. Jets are clustered from PF candidates with the infrared and collinear safe anti-k_{t} algorithm [59], with a size parameter ΔR of 0.5, as implemented in the FastJet package [60]. Corrections to the JES and JER derived from MC simulation are implemented. Charged particles originating from pileup interactions are not included in the jet reconstruction.

The data are compared to simulated events generated either with pythia v6.4.24 Monte Carlo [64] for the QCD processes, or with MadGraph v5.1.3.30 [65, 66], interfaced with pythia v6.4.24 for top ($\text{t}\bar{\text{t}}$ and single-top), Z+jets, W+jets, γ+jets, and diboson (VV) processes. The pythia v6.4.24 program has been set up with a parameter set description for the underlying event referred to as tune Z2* [67, 68]. The generated events are passed through the CMS detector simulation, which is based on geant 4 [69]. Events in simulation are reweighted so that the number of pileup interactions agrees with events in data.

In the studies outlined below, $\text{Z} \to \mu\mu$ events are selected to tune the $\vec{\not{E}}_{\text{T}}$ significance algorithm and to evaluate its performance on events with nominally zero $\vec{\not{E}}_{\text{T}}$. These events are selected to pass a double muon trigger with p_{T} thresholds of 17 GeV and 8 GeV on the leading and sub-leading muons, respectively. Offline, we require exactly two muons satisfying $p_{\text{T}} > 20\,\text{GeV}$ and $|\eta| < 2.1$, lying within the invariant mass window $60 < \text{M}_{\mu\mu} < 120\,\text{GeV}$. Events with more than two muons (isolated or non-isolated) are vetoed to limit the contribution from nonzero-$\vec{\not{E}}_{\text{T}}$ events.

Dijet events provide a sample that also contains nominally zero $\vec{\not{E}}_{\mathrm{T}}$, but test the algorithm at higher jet p_{T} values. These events are required to pass a trigger with a p_{T} threshold of 320 GeV on the leading jet. We also require at least two jets to satisfy $p_{\mathrm{T}} > 200\,\mathrm{GeV}$, with at least one jet satisfying $p_{\mathrm{T}} > 400\,\mathrm{GeV}$.

To evaluate the performance of the algorithm in events with nonzero $\vec{\not{E}}_{\mathrm{T}}$, $\mathrm{W} \to \mathrm{e}\nu$ and semi-leptonic $\mathrm{t\bar{t}}$ samples are employed. To select $\mathrm{W} \to \mathrm{e}\nu$ events, we employ a single-electron trigger with a p_{T} threshold of 27 GeV. Offline, we require one electron satisfying $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.5$. Events with an additional electron satisfying $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$ are vetoed. Semi-leptonic $\mathrm{t\bar{t}}$ events are required to pass one of two triggers: the first requires an electron with $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and a b-tagged jet with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$; the second requires an isolated muon with $p_{\mathrm{T}} > 17\,\mathrm{GeV}$ and a b-tagged jet with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$. Offline, we require at least three jets satisfying $p_{\mathrm{T}} > 45\,\mathrm{GeV}$ with at least two of them b-tagged, at least four jets satisfying $p_{\mathrm{T}} > 20\,\mathrm{GeV}$, and at least one isolated electron or muon. For b-tagging, the CSV algorithm is employed with a tight working point.

5.3 Jet resolutions and tuning

The $\vec{\not{E}}_{\mathrm{T}}$ resolution captured in the covariance matrix $\mathbf{V}$ of Eq. (5.3) is determined mainly by the momentum resolution of the hadronic components of the event. For the purpose of $\vec{\not{E}}_{\mathrm{T}}$ significance we separate the hadronic activity into jets with $p_{\mathrm{T}} \geq 20$ GeV, which are reconstructed with the PF algorithm, and unclustered energy with $p_{\mathrm{T}} < 20$ GeV. The jets are treated as individual objects, each with a unique resolution function depending on the p_{T} and η of the jet, while the objects in the unclustered energy are summed vectorially to produce a single object

with $\vec{p}_{\mathrm{T}} = \sum_i \vec{p}^{\,i}_{\mathrm{T}}$, whose resolution is determined separately. This division separates those components of the event that carry strong azimuthal information and contribute distinctively to the topology of the event from those that are relatively featureless and contribute only to a general broadening of the $\vec{\not{E}}_{\mathrm{T}}$ resolution. Subsequent results are not sensitive to the choice of the 20 GeV threshold.

The resolution functions of hadronic jets are parametrized as a core Gaussian function with additional power-law terms that describe small non-Gaussian tails (Section 4.4.2). The parameter values are determined initially with samples of QCD multijet events generated by PYTHIA v6.4.24 [64], with jets propagated through the full simulation of the CMS detector; the reconstructed and generated values of p_{T}, η, and ϕ are compared to extract resolution shapes.

A full description of a single jet's Gaussian core resolution is given by the covariance matrix,

$$\mathbf{U} = \begin{pmatrix} \sigma^2_{p_{\mathrm{T}}} & 0 \\ 0 & p^2_{\mathrm{T}}\,\sigma^2_{\phi} \end{pmatrix}, \tag{5.5}$$

in which we assume no correlation between p_{T} and ϕ terms. Both $\sigma_{p_{\mathrm{T}}}$ and σ_{ϕ} are functions of p_{T} and η. As written, the covariance matrix $\mathbf{U}$ is in the coordinate system aligned with the jet; in use, all such matrices are rotated by the jet azimuthal angle ϕ into the common CMS xy basis: $\mathbf{V} = \mathbf{R}(\phi)\mathbf{U}\,\mathbf{R}^{-1}(\phi)$.

The widths of the core Gaussian functions obtained from simulation as described above are retuned with data using a $\mathrm{Z} \to \mu\mu$ control sample. The dimuon sample is effectively a zero-$\not{E}_{\mathrm{T}}$ sample and the observed $\vec{\not{E}}_{\mathrm{T}}$ is therefore expected to derive primarily from jet resolution smearing rather than from genuine $\vec{\not{E}}_{\mathrm{T}}$. In this sample, jet activity is modest and the $\vec{\not{E}}_{\mathrm{T}}$ characteristics are dominated by the largely isotropic features of the unclustered energy. The $\vec{\not{E}}_{\mathrm{T}}$ significance therefore

conforms well to the null hypothesis, and we use this fact to optimize the Gaussian widths.

Each Gaussian width, σ^{MC}, obtained from simulation is rescaled by an η-dependent correction factor:

$$\sigma(p_{\mathrm{T}}, \eta) = a(\eta) \times \sigma^{\mathrm{MC}}, \tag{5.6}$$

where the correction factors (in five bins of $|\eta|$) are determined by a likelihood fit over the $\mathrm{Z} \to \mu\mu$ data sample in which we seek to maximize the null hypothesis, $\mathcal{L}(\vec{\varepsilon} = 0)$. This is equivalent to minimizing the quantity

$$-2\log[\mathcal{L}(\vec{\varepsilon} = 0)] = S + \log(\det \mathbf{V}), \tag{5.7}$$

where we have combined Eqs. (5.2) and (5.3). To reduce possible biases stemming from events with sources of genuine $\vec{\not{E}}_{\mathrm{T}}$, the fit is performed iteratively with a restriction to exclude high-significance events.

The best-fit values for the η-dependent tuning parameters can be found in Table 5.1. These five parameters are fit to values near unity, and are intended to provide small corrections to the measured JER values. The fact that $a(\eta)$ fit to values slightly larger than unity is due to, among other factors, differences between JER in data and simulation, differences between the JER in the 7 TeV and 8 TeV datasets collected in 2010 and 2012, respectively, and imperfections in the modeling of the jet and unclustered energy response.

The unclustered energy resolution, σ_{uc}, is parametrized by,

$$\sigma_{uc}^2 \;=\; \sigma_0^2 + \sigma_{\mathrm{s}}^2 \sum_{i=1}^{n} |\vec{p}_{Ti}|, \tag{5.8}$$

where the summation is over the n low-p_{T} objects included in the unclustered energy and σ_0 and σ_{s} are free parameters obtained from the likelihood fit de-

Table 5.1: Best-fit values of $a(\eta)$ parameters introduced in Eq. (5.6).

Parameter	η-bin	Data	Simulation
a_1	[0.0,0.5]	1.15	1.05
a_2	[0.5,1.1]	1.08	0.98
a_3	[1.1,1.7]	1.04	0.96
a_4	[1.7,2.3]	1.13	0.97
a_5	[2.3,5.0]	1.56	1.28

Table 5.2: Best-fit parameters values corresponding to unclustered energy resolution, introduced in Eq. (5.8).

Parameter	Data	Simulation
$\sigma_0(\,\mathrm{GeV})$	0.0	-1.1
$\sigma_\mathrm{s}(\,\mathrm{GeV})^{1/2}$	0.5488	0.5204

scribed above. Because the best fit normally returns $\sigma_0 = 0$ (as one would expect), we see that the resolution of the unclustered energy exhibits the general form $\sigma_\mathrm{uc} \approx \sqrt{n}\,\sigma_X$. Its contribution to the $\vec{\not\!\!E}_\mathrm{T}$ covariance matrix is taken to be isotropic,

$$\mathbf{V}_{uc} = \begin{pmatrix} \sigma_{uc}^2 & 0 \\ 0 & \sigma_{uc}^2 \end{pmatrix} = n\sigma_X^2\,\mathbf{I}, \tag{5.9}$$

as it is constructed from a large number of (mostly) uncorrelated, low-p_T objects. In practice, a slight ellipticity is found in some events, but can be neglected without degrading the $\vec{\not\!\!E}_\mathrm{T}$ significance performance. The quantity σ_X^2 measures the average contribution of low-p_T objects to the $\vec{\not\!\!E}_\mathrm{T}$ covariance. The best-fit values of the σ_0 and σ_s parameters are given in Table: 5.2.

Systematic uncertainties associated with hadronic activity are evaluated using uncertainties on the jet energy scale (2-10%) and the energy scale of low energy

particles entering into the unclustered energy (10%), and are displayed as gray bands in Figs. 5.3-5.6. The systematic uncertainty due to jet energy resolution is captured here as well.

Electron and muon resolutions are assumed to be negligible when compared to those for the hadronic activity in each event, and thus do not enter into the $\vec{\not{E}}_{\mathrm{T}}$ covariance.

5.4 Performance of $\mathcal{S}$

5.4.1 Events with $\not{E}_T = 0$

As $\mathcal{S}$ is χ^2-distributed, an event sample that nominally has no genuine $\vec{\not{E}}_{\mathrm{T}}$ should be flat in the χ^2 probability function for two degrees of freedom, $\mathcal{P}_2(\mathcal{S})$. Here, $\mathcal{P}_2(\mathcal{S})$ is defined such that $1 - \mathcal{P}_2(\mathcal{S})$ is the standard cumulative distribution function of the χ^2 statistic for two degrees of freedom. Both $\mathrm{Z} \to \mu\mu$ and dijet samples from pp collisions are dominated by such events. The dijet sample is defined in Section 5.2; though heavily populated by events with two high-p_{T} jets, it is not restricted by any limit on the maximum number of jets.

We compare the distributions of $\mathcal{S}$ as well as $\mathcal{P}_2(\mathcal{S})$ in data and simulation for both $\mathrm{Z} \to \mu\mu$ and dijet samples in Figs. 5.3 and 5.4. The observed spectrum conforms to a χ^2 distribution in the core region, but begins to slightly deviate from a perfect χ^2 at high values of significance ($\mathcal{S} \gtrsim 9$). Physics backgrounds containing nonzero true $\vec{\not{E}}_{\mathrm{T}}$ (defined here to be $\not{E}_{\mathrm{T}} > 3$ GeV) are present, but are negligible in comparison to the dominant zero-$\not{E}_{\mathrm{T}}$ population. The impact of

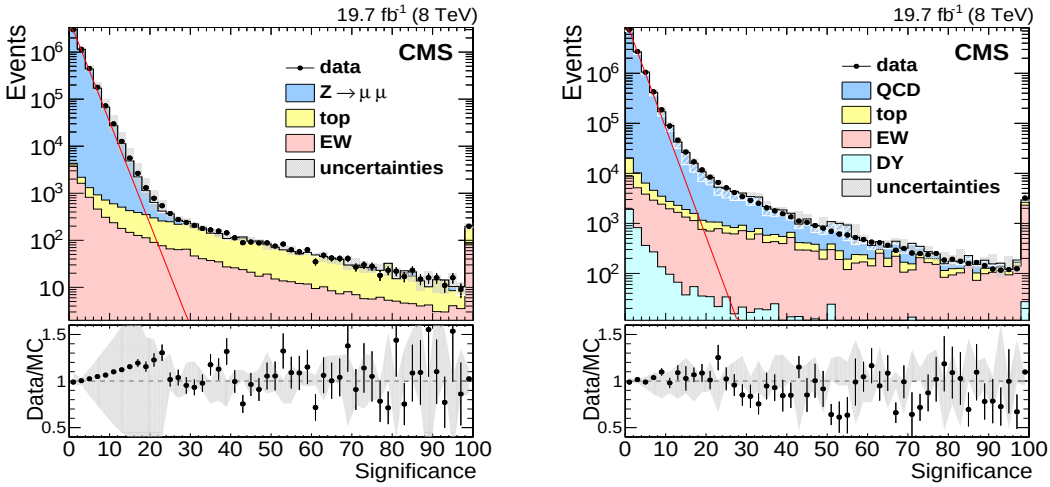

Figure 5.3: Distribution of $\vec{\not{E}}_{\mathrm{T}}$ significance in the (left) $\mathrm{Z} \to \mu\mu$ and (right) dijet samples. The red line corresponds to a χ^2 distribution with two degrees of freedom; the white hatched region shows the distribution of events containing genuine non-zero $\not{E}_{\mathrm{T}}$. The points in the lower panel of each plot show the data/MC ratio, including the statistical uncertainties of both data and simulation; the gray error band displays the systematic uncertainty of the simulation.

$\mathrm{Z} \to \mu\mu$ events with true $\not{E}_{\mathrm{T}}$ due to heavy-quark decays and decays in flight is also found to contribute to the high-$\mathcal{S}$ region. Such events only constitute about 1% of the signal sample in simulated events, however. The general agreement with a χ^2 distribution is also apparent in the $\mathcal{P}_2(\mathcal{S})$ spectra, which are flat over the bulk of events and show an excess at low values of $\mathcal{P}_2(\mathcal{S})$ (high values of $\mathcal{S}$). It is helpful to keep in mind that $\mathcal{P}_2(\mathcal{S}) < 0.01$ corresponds to $\mathcal{S} > 9.2$, $\mathcal{P}_2(\mathcal{S}) < 0.02$ corresponds to $\mathcal{S} > 7.8$, and $\mathcal{P}_2(\mathcal{S}) < 0.05$ corresponds to $\mathcal{S} > 6.0$.

5.4.2 Events with $\not{E}_T \neq 0$

The presence of genuine $\vec{\not{E}}_{\mathrm{T}}$ pushes events to higher values of $\mathcal{S}$ and lower values of $\mathcal{P}_2(\mathcal{S})$, and thus can be used to separate events with genuine $\vec{\not{E}}_{\mathrm{T}}$ from those with

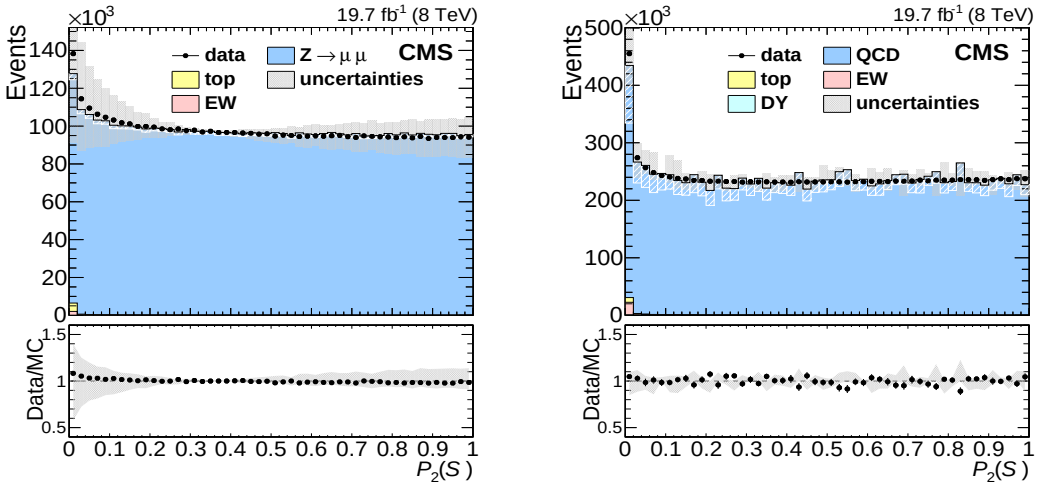

Figure 5.4: Distribution of $\mathcal{P}_2(\mathcal{S})$ in the (left) $Z \to \mu\mu$ and (right) dijet samples. Events that contain a source of genuine $\vec{\not{E}}_T$ are represented by the hatched white region. The points in the lower panel of each plot show the data/MC ratio, including the statistical uncertainties of both data and simulation; the gray error band displays the systematic uncertainty of the simulation.

only resolution-induced $\vec{\not{E}}_T$. To study the discrimination power of the significance variable, we use samples of events containing W boson or $t\bar{t}$ production. The $W \to e\nu$ channel offers a probe of $\vec{\not{E}}_T$ significance in a scenario dominated by genuine $\vec{\not{E}}_T$, accompanied by significant zero-$\vec{\not{E}}_T$ backgrounds; the semileptonic $t\bar{t}$ channel similarly provides a genuine $\vec{\not{E}}_T$ signal, but with background events predominantly from higher-$\not{E}_T$ dileptonic $t\bar{t}$ decays.

The distributions in data and simulation of the $\vec{\not{E}}_T$ significance and corresponding $\mathcal{P}_2(\mathcal{S})$ distributions are shown in Figs. 5.5 and 5.6 for both the $W \to e\nu$ and semi-leptonic $t\bar{t}$ events. Some interesting features are apparent in the composition of simulation events in the significance spectra. In the $W \to e\nu$ channel, events arising from zero true $\vec{\not{E}}_T$ physics channels, such as QCD and Drell–Yan events, are mostly found at low values of significance compared to the broad distribution of non-zero-$\vec{\not{E}}_T$ events. The semi-leptonic $t\bar{t}$ channel has a significant non-zero-$\vec{\not{E}}_T$

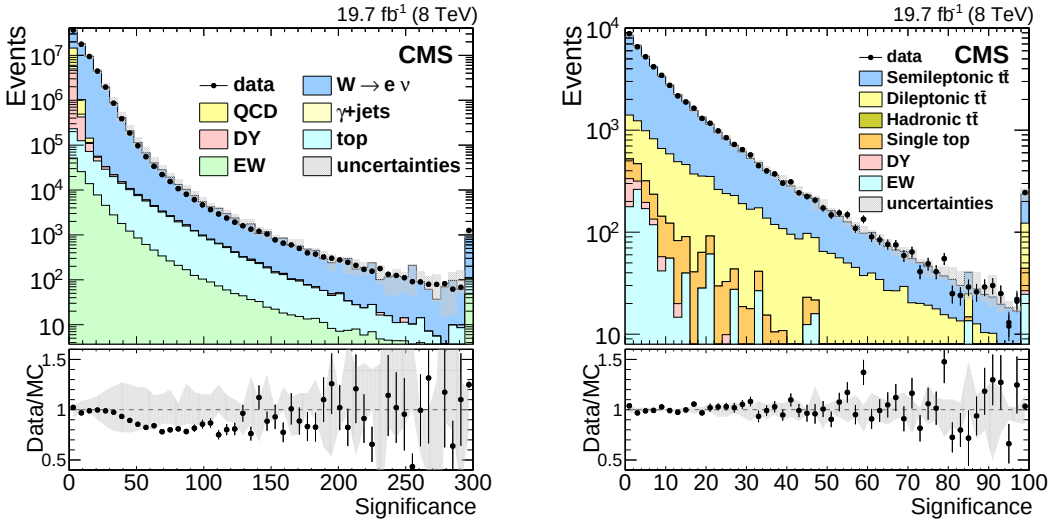

Figure 5.5: Distribution of $\vec{\not{E}}_{\mathrm{T}}$ significance in the (left) $\mathrm{W} \to \mathrm{e}\nu$ and (right) $\mathrm{t\bar{t}}$ events.

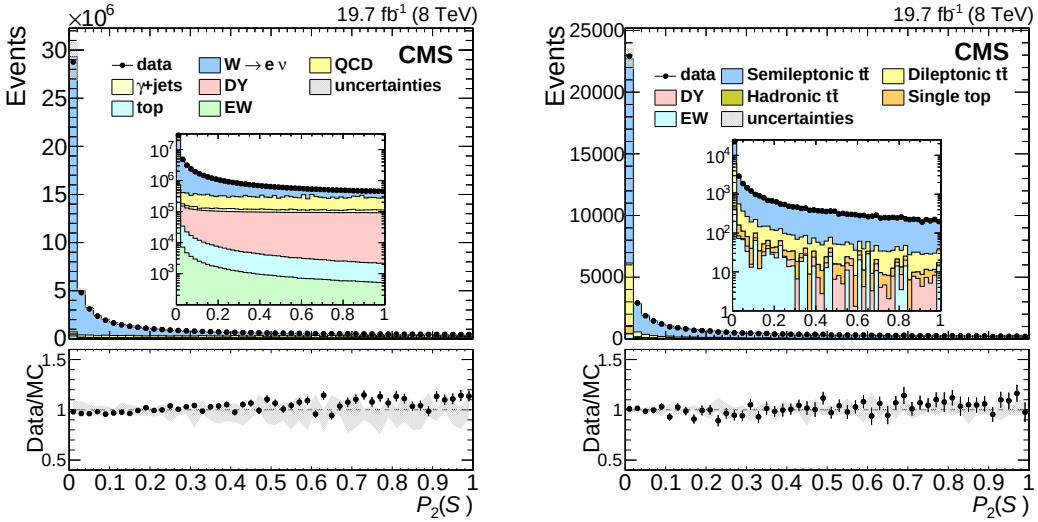

Figure 5.6: Distribution of $\mathcal{P}_2(\mathcal{S})$ in the (left) $\mathrm{W} \to \mathrm{e}\nu$ and (right) $\mathrm{t\bar{t}}$ events. The insets show the same data as the main plots, but with a log scale to show the background components more clearly.

background stemming from dileptonic $\mathrm{t\bar{t}}$ decays. The dileptonic $\mathrm{t\bar{t}}$ spectrum falls more slowly than the semileptonic $\mathrm{t\bar{t}}$ signal in the tail region of $\mathcal{S}$.

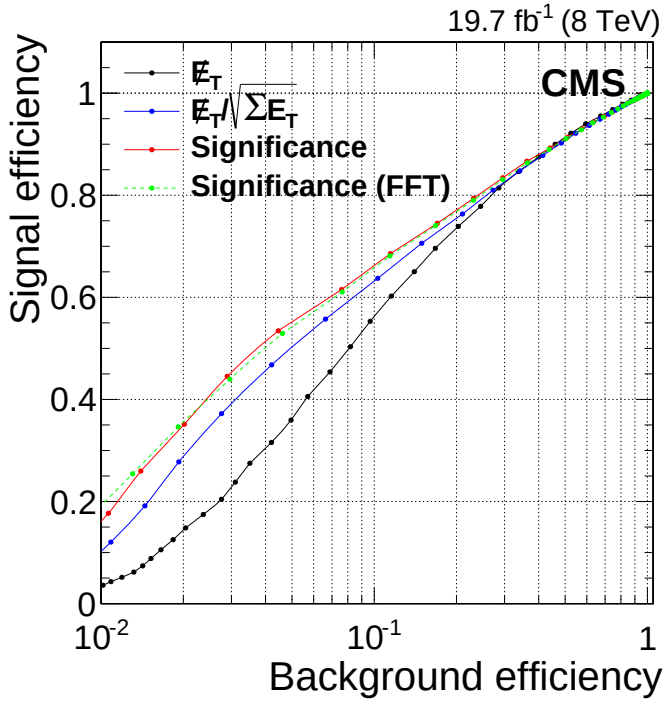

Figure 5.7: Signal versus background efficiencies for W $\to$ eν for various $\vec{\not\!\!E}_{\mathrm{T}}$-based discriminating variables. The FFT significance variable (green dashed line) is discussed in Section 5.6.

5.5 Performance of $\mathcal{S}$ for background rejection

Here we examine the potential gain of introducing the significance variable into the selection criteria for W $\to$ eν and semileptonic t$\bar{\mathrm{t}}$ events. Fig. 5.7 compares the signal and background efficiencies for W $\to$ eν events in simulation, where increasing thresholds are placed on the value of $\mathcal{S}$. (The green curve is discussed in Section 5.6.) In the W $\to$ eν channel, there is a performance benefit in using $\vec{\not\!\!E}_{\mathrm{T}}$ significance when compared to simpler background discrimination variables such as $\not\!\!E_{\mathrm{T}}$ alone or the approximate significance variable $\not\!\!E_{\mathrm{T}}/\sqrt{\sum E_{\mathrm{T}}}$ [70]. For example, choosing a working point with 50% signal efficiency yields a background efficiency of 8.2% using $\not\!\!E_{\mathrm{T}}$, 5.1% using $\not\!\!E_{\mathrm{T}}/\sqrt{\sum E_{\mathrm{T}}}$, and 4.0% using the significance as a discriminating variable. For reference, a 50% signal efficiency working

point corresponds to a $\not\!\!E_{\mathrm{T}} > 40$ GeV requirement. In the semi-leptonic $\mathrm{t\bar{t}}$ channel, $\mathcal{S}$ provides discrimination that is comparable to $\not\!\!E_{\mathrm{T}}$ and $\not\!\!E_{\mathrm{T}}/\sqrt{\sum E_{\mathrm{T}}}$. This reflects the fact that $\mathcal{S}$ is optimized for discriminating events that satisfy the null hypothesis ($\vec{\epsilon} = 0$) from those that do not. In the case of semileptonic $\mathrm{t\bar{t}}$, the dominant background contribution comes from dileptonic $\mathrm{t\bar{t}}$ decays with large, genuine $\not\!\!E_{\mathrm{T}}$.

We have also evaluated the performance benefit of modeling individual jet resolutions down to 3 GeV, as in Ref. [61], as an alternative to the current threshold of 20 GeV. Using a lower threshold for individual jets can potentially provide more detailed information about the low-p_{T} hadronic activity, but we find that the performance in the $\mathrm{W} \to \mathrm{e}\nu$ channel is essentially indistinguishable when implemented with these two different thresholds, and therefore use the simpler 20 GeV threshold.

5.5.1 Pileup

The $\vec{\not\!\!E}_{\mathrm{T}}$ significance variable exhibits simple behavior as a function of the number of pileup interactions. For event samples such as the $\mathrm{Z} \to \mu\mu$ and dijet selections, in which in most events there is no source of true $\not\!\!E_{\mathrm{T}}$, the $\mathcal{S}$ value remains essentially constant as the number of primary vertices increases. In samples such as $\mathrm{W} \to \mathrm{e}\nu$ and $\mathrm{t\bar{t}}$, where the average value of $\not\!\!E_{\mathrm{T}}$ is non-zero, a decrease with increasing pileup is seen. This behavior can be derived formally from the expression for $\mathcal{S}$ given in Eq. (5.3) with the isotropic model of unclustered energy given in Eq. (5.9) if the additional covariance due to n pileup vertices is incorporated via the replacement $\mathbf{V} \to \mathbf{V}_0 + n\sigma^2\mathbf{I}$. In this transformation $\mathbf{V}_0$ represents the covariance matrix in the absence of pileup. It is also confirmed empirically in Fig. 5.8. As a side point, we note that $\langle\mathcal{S}\rangle \approx 2$ for the zero-$\not\!\!E_{\mathrm{T}}$ events, as one expects for a χ^2 variable with

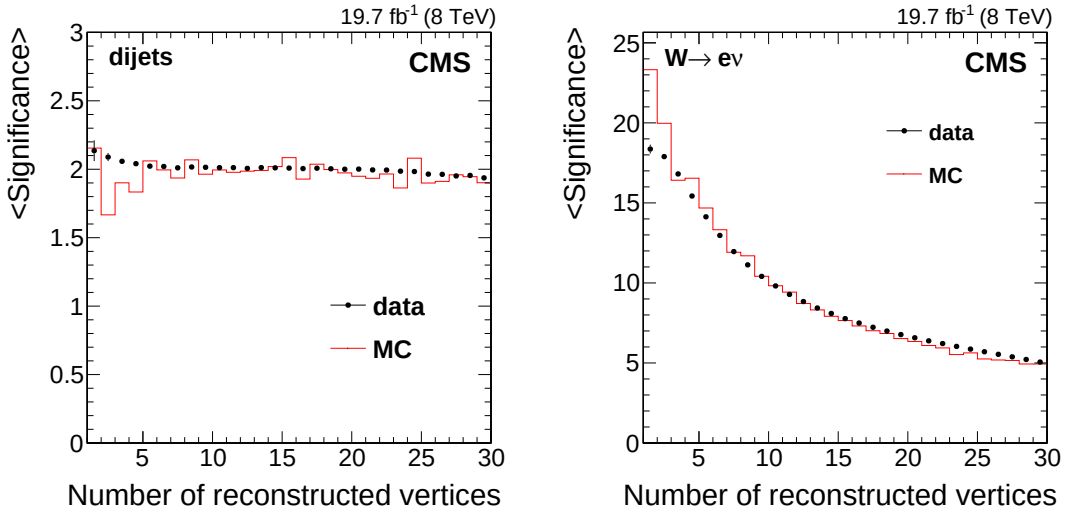

Figure 5.8: The average $\vec{\not\!\!E}_{\mathrm{T}}$ significance versus the number of reconstructed vertices for (left) dijet and (right) $\mathrm{W} \to \mathrm{e}\nu$ event samples.

two degrees of freedom.

As a result of the pileup dependence observed for genuine $\not\!\!E_{\mathrm{T}}$ events, the background rejection performance of the $\vec{\not\!\!E}_{\mathrm{T}}$ significance can also exhibit a dependence on pileup. This is demonstrated for the $\mathrm{W} \to \mathrm{e}\nu$ channel in Fig. 5.9. Here we see a decreasing signal efficiency as the pileup increases. It is also apparent that while the efficiencies of non-zero-$\not\!\!E_{\mathrm{T}}$ signal events depend on pileup, the efficiencies for the zero-$\not\!\!E_{\mathrm{T}}$ background events are relatively stable.

5.6 Treatment of non-Gaussian resolutions

As noted earlier, the jet p_{T} resolution functions exhibit non-Gaussian tails. The challenge presented by such tails lies in the convolution integrals needed to compute the $\not\!\!E_{\mathrm{T}}$ likelihood function. These can be done analytically for Gaussian resolutions, but not when non-Gaussian elements are introduced and direct, nu-

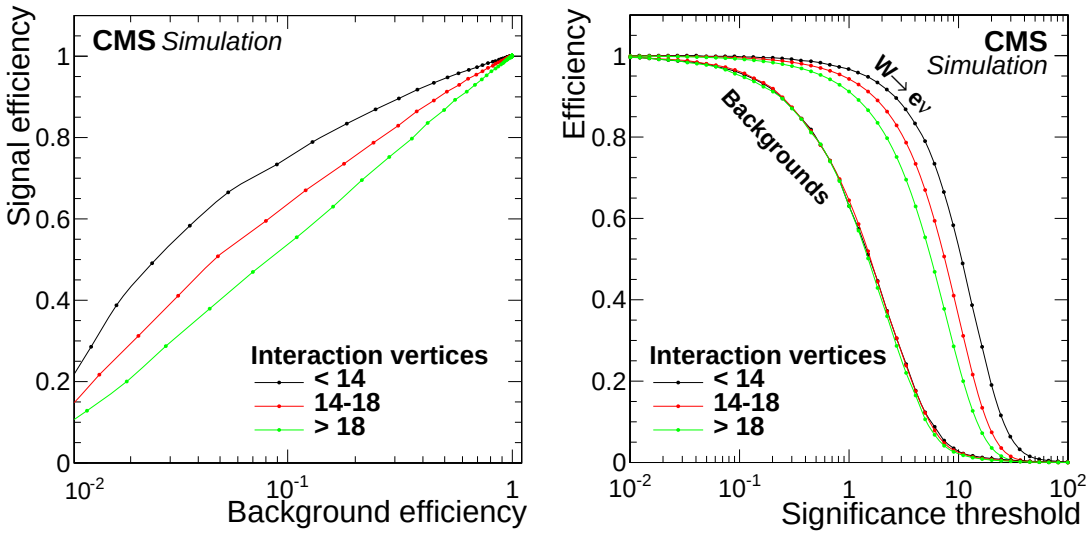

Figure 5.9: Efficiency curves for $\vec{\not{E}}_{\mathrm{T}}$ significance in W $\to$ eν channel in three regions defined by the number of reconstructed vertices. The signal versus background efficiencies are shown in the left pane. In the right pane, the signal (right) and background (left) efficiencies are shown separately as a function of the threshold on $\mathcal{S}$.

merical convolution is prohibitively slow. The convolution process, however, can be reduced under Fourier transformation to simple multiplication of the transformed functions. With this approach, each jet resolution function $R_i(p_x, p_y)$ is transformed to $\widetilde{R}_i(k_x, k_y)$, and then the product $\prod_{i=1}^{n} \widetilde{R}_i(k_x, k_y)$ is computed and back-transformed to yield the fully convolved result. When computed with fast Fourier transform (FFT) techniques, this method enables the required convolutions to be done at a speed that, while slower than the evaluation of analytic functions, is still well within reason for late stages of analysis. Both R and $\widetilde{R}$ are discretized on 2-dimensional grids in their respective spaces, and the resulting discretized likelihood function is smoothed by cubic spline interpolation before computing the significance. Care is taken in defining the grids to avoid artifacts that can result from aliasing. To verify the validity of this FFT method and its implementation, we have compared the results of the FFT and analytic methods

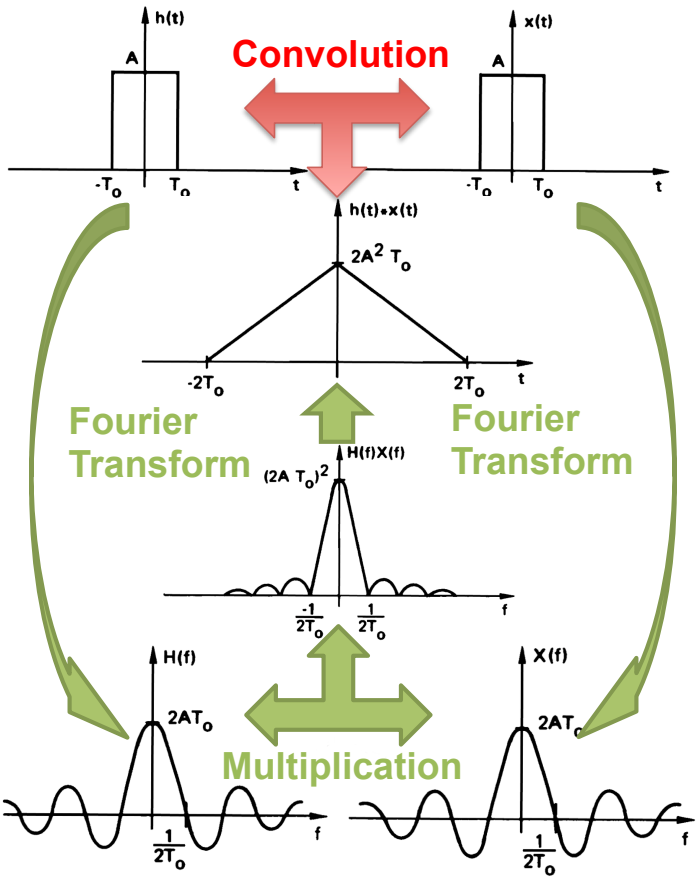

Figure 5.10: Flow chart of the Fast Fourier Transform (FFT) technique to convolve non-Gaussian resolutions [27].

for cases where only Gaussian resolutions are used and find the two methods yield identical results.

To demonstrate the potential utility of the non-Gaussian treatment, we compare $\vec{\not{E}}_{\mathrm{T}}$ significance computed with the FFT and with the analytic method. For the comparison, we use the dijet event sample, as there is sufficient high-p_{T} hadronic activity to exhibit clearly the effects of non-Gaussian contributions to the resolution. Fig. 5.11 shows the results of the comparison. The significance distribution is plotted in the left panel, with the black histogram computed

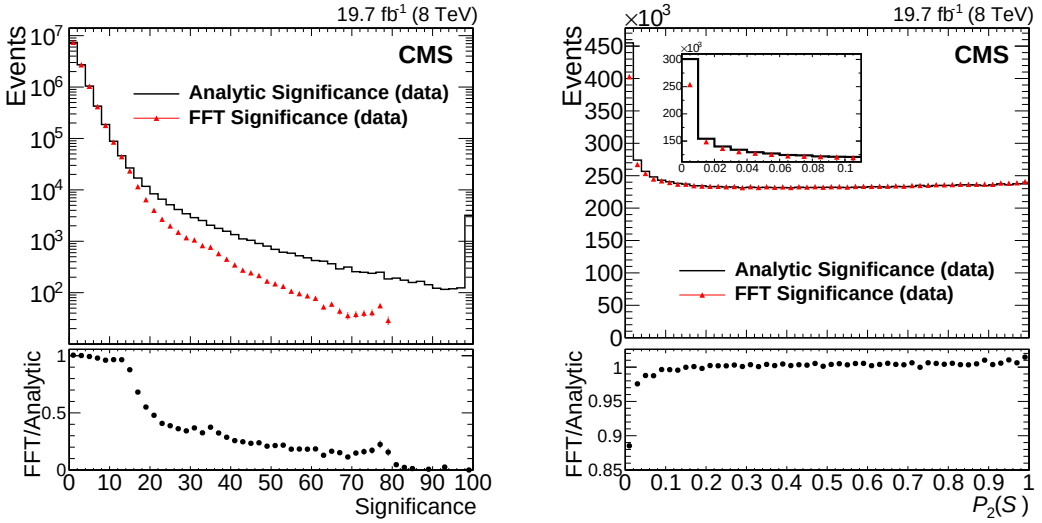

Figure 5.11: Comparisons in dijet events of the FFT (non-Gaussian) and analytic (Gaussian) methods for calculating $\vec{\not\!\!E}_{\mathrm{T}}$ significance. Left: $\vec{\not\!\!E}_{\mathrm{T}}$ significance distribution. Right: $\mathcal{P}_2(\mathcal{S})$ distribution. For this figure, both the FFT (red triangles) and analytic (black histogram) algorithms are applied only to data. The analogous MC distributions for the analytic method are shown in Figs. 5.3 and 5.4. Non-Gaussian significance values of $\mathcal{S} \gtrsim 80$ are suppressed due to the finite number of significant figures available to double precision variables used in the FFT algorithm.

by the analytic method (i.e. assuming only Gaussian resolutions), and red data points computed with the FFT algorithm (using full resolution functions). The steeper fall of the red points demonstrates that the FFT algorithm helps to reduce the excess of high-significance values that arise in the analytic method where jet measurement uncertainty is underestimated by the Gaussian approximation. The right-hand panel shows the corresponding reduction of events in the lowest bin of the $\mathcal{P}_2(\mathcal{S})$ distribution. The remaining excess in that bin is partly due to events with genuine $\not\!\!E_{\mathrm{T}}$ that arise from semileptonic decays of hadrons. After taking into account these genuine $\not\!\!E_{\mathrm{T}}$ components and other extraneous backgrounds from $\mathrm{t\bar{t}}$ and vector boson production, the net impact of the FFT algorithm is to reduce the

excess of zero-$\not\!\!E_{\mathrm{T}}$ events in the high-significance, low-$\mathcal{P}_2(\mathcal{S})$ bin ($\mathcal{P}_2(\mathcal{S}) < 0.02$) by a factor of two. Removal of the remaining zero-$\not\!\!E_{\mathrm{T}}$ events in this bin will require deeper understanding of the jet-by-jet resolution variations that are not captured by the average parametrizations currently available.

5.7 Commissioning for Run 2

The $\vec{\not\!\!E}_{\mathrm{T}}$ significance algorithm was retuned and tested using 13 TeV proton-proton collision data collected during 2016, corresponding to an integrated luminosity of $12.9\,\mathrm{fb}^{-1}$. Events containing Z boson decays, $\mathrm{Z} \to \mu\mu$ and $\mathrm{Z} \to \mathrm{ee}$, are selected using dilepton triggers with p_{T}-thresholds of 17/8 GeV for muons and 23/12 GeV for electrons, and single lepton triggers with higher thresholds of 24 GeV and 27 GeV for muons and electrons, respectively. The dilepton pair is required to be contained in the mass window $81\,\mathrm{GeV} < \mathrm{M}_{\ell\ell} < 101\,\mathrm{GeV}$. Offline, muons are required to satisfy $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$, and electrons are required to satisfy $p_{\mathrm{T}} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$. A requirement of at least two jets with $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ is added to improve agreement between data and simulation in these samples. A complete overview of MC samples and event selection criteria is documented in Ref. [62].

Tuning of the $\vec{\not\!\!E}_{\mathrm{T}}$ algorithm is conducted using the $\mathrm{Z} \to \mu\mu$ event selection. The values of σ^{MC} are updated with new measurements of the jet response based on the 13 TeV dataset. Scale factors are applied to the values of σ^{MC} in data to account for measured differences in the jet response between data and simulation. In addition, the p_{T} threshold separating jets from the unclustered energy is decreased from 20 GeV to 15 GeV.

The η-bin definitions are updated to correspond to the updated scale factors.

Table 5.3: Values of $a(\eta)$ parameters introduced in Eq. (5.6).

Parameter	η-bin	Data	Simulation
a_1	[0.0,0.8]	1.26	1.29
a_2	[0.8,1.3]	1.14	1.19
a_3	[1.3,1.9]	1.13	1.07
a_4	[1.9,2.5]	1.13	1.13
a_5	[2.5,5.0]	1.06	1.12

Table 5.4: Parameters values corresponding to unclustered energy resolution, introduced in Eq. (5.8).

Parameter	Data	Simulation
$\sigma_0(\,\mathrm{GeV})$	3.3	0.0
$\sigma_\mathrm{s}(\,\mathrm{GeV})^{1/2}$	0.5961	0.6504

The tuning parameters $a(\eta)$ corresponding to the jet sector are given in Table 5.3. The parameters corresponding to the unclustered energy are given in Table 5.4. The updated parameter values show an increase when compared to the 8 TeV tuning studies, especially for small values of jet $|\eta|$. In addition, the agreement in the unclustered energy parameter values between data and simulation has worsened. These features point to possible differences between the 13 TeV and 8 TeV datasets that need further investigation. Distributions of $\vec{\not{E}}_\mathrm{T}$ significance for events containing Z-boson decays are shown in Fig. 5.12.

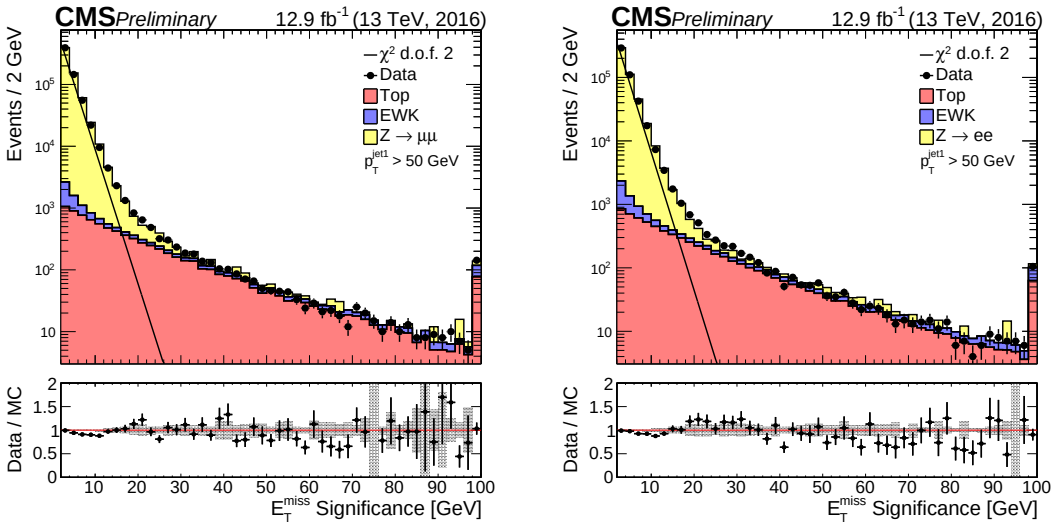

Figure 5.12: Distribution of $\vec{\not{E}}_{\mathrm{T}}$ significance in the Z $\rightarrow$ $\mu\mu$ and Z $\rightarrow$ ee channel for one jet selection, respectively. The black straight line corresponds to a χ^2 distribution with two degrees of freedom. The grey band in the lower pad displays the statistical uncertainty on the simulation

CHAPTER 6

GAUSSIAN PROCESSES

No more training, do you require.

– Yoda

The measurement of the top quark mass described in Chapter 7 requires a precise modeling of distribution shapes. Each distribution has the form $f(x|\mathrm{M_t}, \mathrm{JSF})$, where x is the value of an observable ($\mathrm{M_{b\ell}}$, $\mathrm{M_{T2}^{bb}}$, or $\mathrm{M_{b\ell\nu}}$), and the free parameters $\mathrm{M_t}$ and JSF are to be determined in the event-by-event likelihood fit described in Section 7.4. A central feature of event-by-event likelihoods, such as the one in Eq. (7.7), is the requirement of a smooth function $f(x|\mathrm{M_t}, \mathrm{JSF})$ that can be evaluated at any point x on its range. In addition, the function's evolution in the parameters $\mathrm{M_t}$ and JSF must be understood, so that it can be smoothly varied by a numerical minimization algorithm to determine the best-fit parameter values. A variety of techniques have been proposed for modeling distribution shapes in high energy physics analysis. A common approach uses parametric functions, where the function parameters themselves carry a dependence on the fit parameter of interest (e.g. $\mathrm{M_t}$ or JSF). Although such methods have the advantage of being fairly simple and easy to implement, they often introduce unwanted bias that must be calibrated away at late stages of the analysis.

In this chapter, we outline a Gaussian Process (GP) regression technique that has two main advantages over other commonly-used shape estimation methods. First, the GP shape is nonparametric, determined only by a set of training points and hyperparameters that regulate smoothing; and second, it can be easily trained as a function of several variables simultaneously. The latter feature allows one to

capture the smooth evolution of the relevant distribution shapes as the M_t and JSF parameters are varied. A formal introduction to GPs can be found in Refs. [71, 72]. Here, we give an overview of topics relevant to the top quark mass measurement in Chapter 7.

6.1 An informal introduction

To complement the formal descriptions in Refs. [71, 72], we provide a discussion of several core concepts underlying GP regression without striving for absolute mathematical rigor. We begin with a function f, where the value of the function at any point $\mathbf{u}_i$ is given by $f(\mathbf{u}_i)$. (The point $\mathbf{u}_i$ may be a vector with an arbitrary number of dimensions.) Given a collection of N such points, $\{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$, we can define a column vector, $\mathcal{G}$, of corresponding function values:

$$\mathcal{G} = \begin{bmatrix} f(\mathbf{u}_1) \\ f(\mathbf{u}_2) \\ \vdots \\ f(\mathbf{u}_N) \end{bmatrix}. \tag{6.1}$$

In a GP, the values $\{f(\mathbf{u}_1), \ldots, f(\mathbf{u}_N)\}$ are related by a multivariate Gaussian distribution. To illustrate this, we consider samples $\mathcal{G}_a$ and $\mathcal{G}_b$ with $N = 50$ points from two different GPs, shown in Fig. 6.1. In $\mathcal{G}_a$ on the left, each value $f(\mathbf{u}_i)$ is drawn at random from a one-dimensional Gaussian distribution – the value at each point $\mathbf{u}_i$ is independent from its neighboring value at $\mathbf{u}_{i+1}$. In $\mathcal{G}_b$ on the right, each value is correlated with its neighbors, with a covariance between any two values $f(\mathbf{u}_i)$ and $f(\mathbf{u}_j)$ defined by:

$$\mathrm{cov}(f(\mathbf{u}_i), f(\mathbf{u}_j)) = \theta_0 \exp\left\{-\frac{1}{2\theta_1^2}(\mathbf{u}_i - \mathbf{u}_j)^2\right\}. \tag{6.2}$$

Here θ_0 and θ_1 are hyperparameters that control the properties of the shape. In this case, the correlation between two values is a function of their separation, $|\mathbf{u}_i - \mathbf{u}_j|$. This has the effect of producing the smooth shape at the right, in contrast to the noisy distribution on the left.

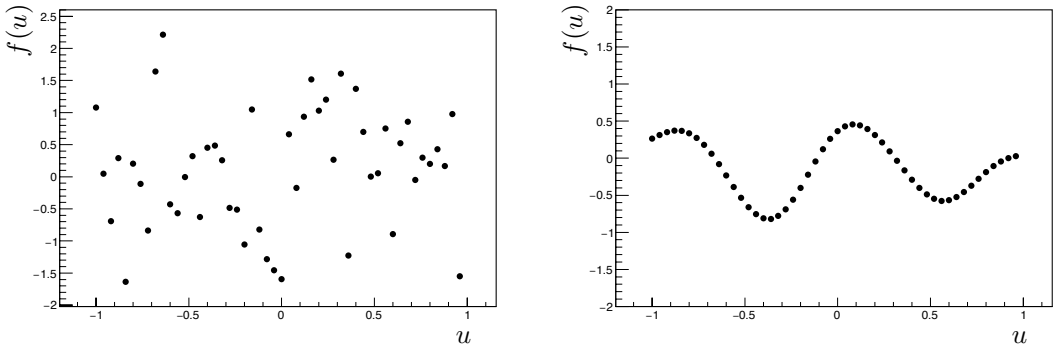

Figure 6.1: Sample with $N = 50$ points drawn from two different GPs. In $\mathcal{G}_a$ (left), each value $f(\mathbf{u}_i)$ is drawn independently from a one-dimensional Gaussian distribution. This corresponds to a constant GP kernel. In $\mathcal{G}_b$ (right), the entire vector is drawn at random from an N-dimensional Gaussian distribution with the Gaussian GP kernel defined in Eq. (6.2), where $\theta_1 = 0.2$.

Both samples in Fig. 6.1 follow a multivariate Gaussian distribution of the form:

$$\mathbf{G}(\mathbf{0}, \mathbf{C}_N) \sim \exp\left(-\frac{1}{2}\mathcal{G}^{\mathrm{T}}\mathbf{C}_N^{-1}\mathcal{G}\right), \tag{6.3}$$

where $\mathbf{C}_N$ is an $N \times N$ covariance matrix. In the sample $\mathcal{G}_a$ on the left, $\mathbf{C}_N = C \cdot \mathbb{I}_N$ is a diagonal matrix, and the distribution can be factorized into a product of independent one-dimensional Gaussian distributions. Hence, each value $f(\mathbf{u}_i)$ can be drawn independently. In the sample $\mathcal{G}_b$ on the right, $\mathbf{C}_N$ is determined by the covariance function, or kernel, in Eq. (6.2). It contains non-diagonal terms that correlate neighboring points on the shape – any two values $f(\mathbf{u}_i)$ and $f(\mathbf{u}_j)$ are likely to be close together if $\mathbf{u}_i$ and $\mathbf{u}_j$ are close together. To generate the shape at the right, the function values $f(\mathbf{u}_i)$ are obtained by drawing the entire

vector $\mathcal{G}_b$ at random from the distribution $\mathbf{G}(\mathbf{0}, \mathbf{C}_N)$ defined in Eq. (6.3).[1] The hyperparameters θ_0 and θ_1 that define the covariance in Eq. (6.2) have an impact on the properties of $\mathcal{G}$. In the figure, $\theta_1 = 0.2$ is chosen, setting a characteristic lengthscale for fluctuations of the shape.

The GP samples $\mathcal{G}_a$ and $\mathcal{G}_b$ in Fig. 6.1 use a constant kernel and a Gaussian kernel, respectively. In general, the GP kernel can be any function of two variables $\mathbf{u}_i$ and $\mathbf{u}_j$ that is positive semidefinite.[2] Fig. 6.2 shows several additional GP samples that demonstrate properties specific to their kernels. The effect of the lengthscale parameter θ_1 in the Gaussian kernel from Eq. (6.2) is evident by comparing the samples on the top left and top right. Samples obtained using an Ornstein-Uhlenbeck kernel, $\mathrm{cov}(f(\mathbf{u}_i), f(\mathbf{u}_j)) = \theta_0 \exp\left\{-\frac{1}{2\theta_1^2}|\mathbf{u}_i - \mathbf{u}_j|\right\}$, and a polynomial kernel ($\mathrm{cov}(f(\mathbf{u}_i), f(\mathbf{u}_j)) = \mathbf{u}_i\mathbf{u}_j$) are shown on the bottom right and bottom left, respectively. The Gaussian kernel results in a smooth shape with variations at well-defined intervals, and is often used for regression, as discussed in Section 6.2 below.

6.2 GPs for regression

In Section 6.1, we have shown that GP samples can be drawn at random from a multivariate Gaussian distribution with covariance $\mathbf{C}_N$. In regression, we are given a set of *training points*, $\{\mathbf{u}_1, \ldots, \mathbf{u}_N\}$, at which the values $\{f(\mathbf{u}_1), \ldots, f(\mathbf{u}_N)\}$ are

[1] A random vector $\mathcal{G}$ can be drawn from an N-dimensional Gaussian distribution with $N \times N$ covariance matrix $\mathbf{C}_N$ and mean $\mathbf{m}$ using a scalar Gaussian random number generator. First, the Cholesky decomposition $\mathbf{C}_N = LL^{\mathrm{T}}$ is computed, where L is a lower triangular matrix. Then, N random values $\{x_1, \ldots, x_N\}$ are drawn from a unit normal distribution by multiple calls to a random number generator. The random vector $\mathcal{G}$ is then given by: $\mathcal{G} = \mathbf{m} + L\mathbf{x}$, which will have the desired distribution with mean $\mathbf{m}$ and covariance $\mathbf{C}_N$.

[2] The GP kernel with covariance matrix $\mathbf{C}_N$ is positive semidefinite if for every non-zero column vector Z of N real numbers, the scalar $Z^{\mathrm{T}}\mathbf{C}_N Z$ is non-negative.

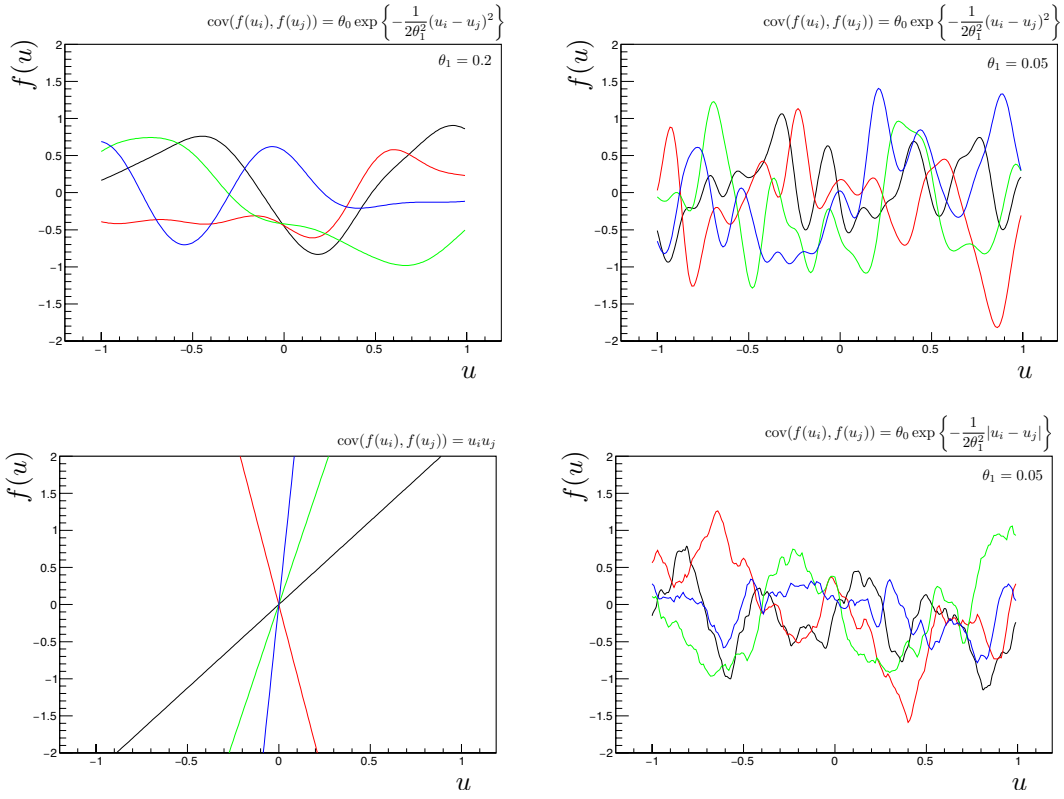

Figure 6.2: GP samples drawn at random with $N = 200$ points. The resulting shapes are interpolated for aesthetic clarity. In the top row, a Gaussian kernel (Eq. (6.2)) is used. The length-scale parameter is set to (top left) $\theta_1 = 0.2$ and (top right) $\theta_1 = 0.05$. The bottom row contains samples using non-Gaussian kernels, given by: (left) $\mathrm{cov}(f(\mathbf{u}_i), f(\mathbf{u}_j)) = \mathbf{u}_i \mathbf{u}_j$ and (right) $\mathrm{cov}(f(\mathbf{u}_i), f(\mathbf{u}_j)) = \theta_0 \exp\left\{-\frac{1}{2\theta_1^2}|\mathbf{u}_i - \mathbf{u}_j|\right\}$ with $\theta_1 = 0.05$. The latter is known as the Ornstein-Uhlenbeck kernel, which describes a form of Brownian motion.

known. We wish to find a function that passes through the training point values and interpolates smoothly between them. More specifically, the function is to be evaluated at any *test point*, $\mathbf{u}_{N+1}$, at which the value of $f(\mathbf{u}_{N+1})$ is not known *a priori*.

The general process of GP regression is illustrated in Fig. 6.3. In Step 1, we draw sample functions from a GP prior (Eq. (6.3)) defined by a kernel of our

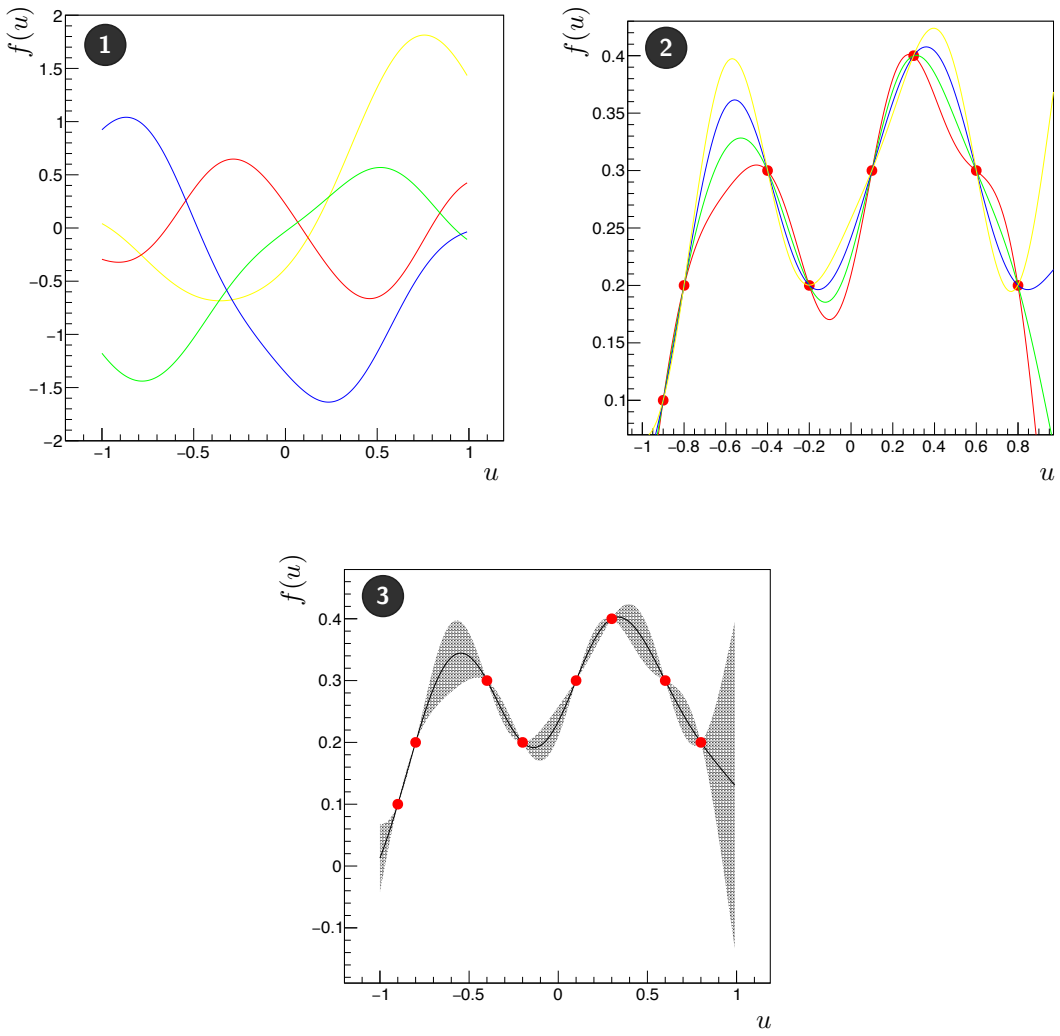

Figure 6.3: Illustration of GP regression in three steps. (Left) functions are drawn at random from a GP prior. (Center) functions are drawn at random from the GP posterior, determined by the training points in red. (Right) the mean (black line) and standard deviation (shaded area) of all possible posterior functions are shown. The mean is taken to be the predictive GP function, with the standard deviation corresponding to uncertainties in the regression.

choosing. In Step 2, instead of drawing sample functions at random from the GP prior, we choose only those that pass through the training points shown in red, i.e. we select them from the GP *posterior* that incorporates our knowledge of the training points. In Step 3, we compute the mean and standard deviation of all possible functions consistent with the training points. The mean (black line) is taken to be the predictive GP function corresponding to the known training points. The standard deviation (shaded band) is a measure of uncertainty in the regression – it is small near the training points, where the values of the shape are well known, and large elsewhere.

In practice, the steps described above are carried out by employing several properties of the multivariate Gaussian distribution. The GP vector in Eq. (6.1) is first populated with entries, $\{f(\mathbf{u}_1), \ldots, f(\mathbf{u}_N), f(\mathbf{u}_{N+1})\}$, corresponding to the N training points and one test point $\mathbf{u}_{N+1}$. These values are related by the multivariate Gaussian distribution in Eq. (6.3). A cross-sectional view of this distribution is represented by the red contour in Fig. 6.4. Given this distribution, we impose our knowledge of the N training points by fixing the distribution to each of their values, indicated by the blue square. The possible values of $f(\mathbf{u}_{\text{test}})$ are now constrained to lie along the blue line, giving rise to the *conditional* Gaussian distribution indicated by the blue dashed curve. The mean of the conditional Gaussian (green circle) is taken to be the value of the GP shape at the test point, and its width is taken to be the modeling uncertainty.

Figs. 6.3 and 6.4 provide two different perspectives on the same underlying process. The red contour in Fig. 6.4 corresponds to the sample functions in Fig. 6.3, Step 1. These functions are sampled from the prior Gaussian distribution with zero mean, just as the contour is centered at the origin. The conditioning processes

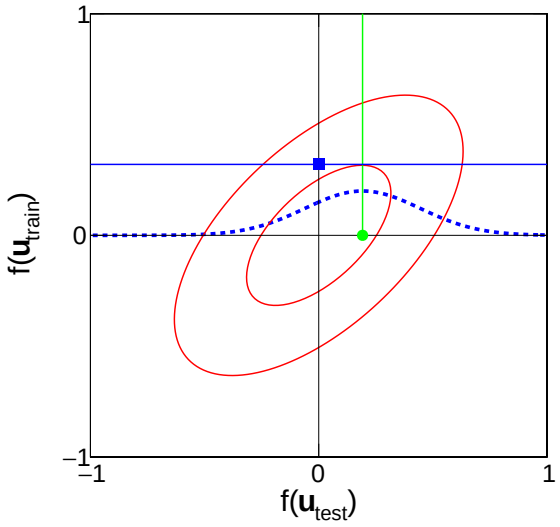

Figure 6.4: Demonstration of the GP *conditioning* process (Eqs. (6.4) and (6.5)) for one training point, $\mathbf{u}_{\text{train}}$, and one test point, $\mathbf{u}_{\text{test}}$. The covariance between the corresponding function values is represented by the red ellipse. The known value $f(\mathbf{u}_{\text{train}})$ (blue square) determines the mean value of $f(\mathbf{u}_{\text{test}})$ (green circle).

represented by the blue solid line corresponds to Step 2, where sample functions drawn from the posterior distribution are constrained to the known training points. Finally, the mean and width of $f(\mathbf{u}_{\text{test}})$ represented by the green circle and blue dashed line correspond to the predictive GP function in Step 3.

The conditioning process described above can be reduced to a set of matrix operations. The mean, μ_{N+1}, and variance, σ^2_{N+1}, of the predictive GP function $f(\mathbf{u})$ at test point $\mathbf{u}_{N+1}$ are given by:

$$\mu_{N+1} = \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{t}, \tag{6.4}$$

$$\sigma^2_{N+1} = c - \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{k}, \tag{6.5}$$

where $\mathbf{t}$ is a column vector containing the $f(\mathbf{u}_i)$ values for all N training points, $\mathbf{k} = \text{cov}(f(\mathbf{u}_i), f(\mathbf{u}_{N+1}))$ is the covariance between the value of f at the i^{th} training point and the value at the test point, and $c = \text{cov}(f(\mathbf{u}_{N+1}), f(\mathbf{u}_{N+1}))$ is the variance

of f at the test point. This result is obtained by defining a $(N+1) \times (N+1)$ covariance matrix corresponding to the N training points and one test point, and partitioning it so that:

$$\mathbf{C}_{N+1} = \begin{bmatrix} \mathbf{C}_N & \mathbf{k} \\ \mathbf{k} & c \end{bmatrix}, \tag{6.6}$$

where the covariance matrix $\mathbf{C}_N$ corresponds to the training points. Then, the mean and variance of the conditional Gaussian distribution (Eqs. (6.4) and (6.5)) can be written in terms of $\mathbf{C}_N$ and the vector $\mathbf{t}$ containing the known training point values.

In summary, the 'Gaussian' in GP can refer to the distribution of possible values of the predictive function f in a regression scenario. The value at a single point, $f(\mathbf{u}_i)$, is distributed according to a one-dimensional Gaussian rather than being treated as an exact quantity. The mean of this Gaussian is the most probable value of the shape at that point $\mathbf{u}_i$, and it is the value used for likelihood fitting (Section 7.4); the variance stems from the modeling uncertainty inherent in the GP regression process. The values $f(\mathbf{u}_i)$ and $f(\mathbf{u}_j)$ at any two points follow a two-dimensional Gaussian distribution and are related by a covariance. The correlation between $f(\mathbf{u}_i)$ and $f(\mathbf{u}_j)$ determines the degree to which the GP shape is allowed to vary between the points $\mathbf{u}_i$ and $\mathbf{u}_j$. By extension, any N values of the shape are described by an N-dimensional Gaussian distribution, and are related by an $N \times N$ covariance matrix. To determine the value of the shape at a test point $\mathbf{u}_{N+1}$, an $N+1$-dimensional Gaussian distribution is constructed relating the training point values $f(\mathbf{u}_1) \ldots f(\mathbf{u}_N)$ to the test point value $f(\mathbf{u}_{N+1})$. Then, $f(\mathbf{u}_1) \ldots f(\mathbf{u}_N)$ are fixed to their known values, and the $N+1$-dimensional Gaussian distribution is reduced to a one-dimensional *conditional* Gaussian distribution representing the possible values of $f(\mathbf{u}_{N+1})$.

6.3 GPs for top quark mass measurement

The likelihood fit described in Section 7.4 uses distribution shapes of the form $f(x|\mathrm{M_t}, \mathrm{JSF})$, where x is the value of an observable ($\mathrm{M_{b\ell}}$, $\mathrm{M_{T2}^{bb}}$, or $\mathrm{M_{b\ell\nu}}$), and $\mathrm{M_t}$ and JSF are free parameters in the fit. The shapes f are shown in Figs. 7.2-7.5 for each observable, where the free parameters are set to $\mathrm{M_t} = 166.5, 172.5$, or $178.5\,\mathrm{GeV}$ and JSF = 1. In the figure, these shapes are represented as functions of a single variable (the observable x) with $\mathrm{M_t}$ and JSF fixed. In GP regression, however, each shape is treated as a function of all three quantities, and can be described as a probability density in three dimensions.

For the following discussion, we define a point, $\mathbf{u}_i$, on each GP shape by its position in x, $\mathrm{M_t}$, and JSF:

$$\mathbf{u}_i \equiv (x_i, \mathrm{M_t}_i, \mathrm{JSF}_i). \tag{6.7}$$

The value of the shape at $\mathbf{u}_i$ is given by $f(\mathbf{u}_i) = f(x_i|\mathrm{M_t}_i, \mathrm{JSF}_i)$. The point $\mathbf{u}_i$ can be a training point, at which the value of f is known and used as an input into the GP regression process; or it can be a test point, at which the value of f is to be determined. Each GP shape is trained using binned distributions of the observable x in MC simulation. For each observable, 35 binned distributions are used, corresponding to seven values of $\mathrm{M_t^{MC}}$ ranging from 166.5 to 178.5 GeV and five values of JSF ranging from 0.97 to 1.03. Each distribution has 75 bins in x, yielding a total of 2625 training points. The GP regression technique interpolates between the discrete values of x, $\mathrm{M_t}$, and JSF covered by these training points to provide a shape that is smooth over its range.

The covariance $\mathrm{cov}(f(\mathbf{u}_i), f(\mathbf{u}_j))$ between any two points is determined by a kernel function similar to the Gaussian kernel described in Eq. (6.2). The choice

of a Gaussian kernel ensures that the correlation between any two values of the distribution is suppressed at a large separation. In practice, the kernel is a three-dimensional function which controls the smoothness of the shape along x, $\mathrm{M_t}$, and JSF. It also includes a correlation term between $\mathrm{M_t}$ and JSF to reflect the kinematic relationship between them. The result is a product of a one-dimensional Gaussian (controlling the smoothness along x) with a two-dimensional Gaussian (controlling the smoothness along $\mathrm{M_t}$ and JSF). For any two points $\mathbf{u}_i$ and $\mathbf{u}_j$ on the shape, the kernel is given by:

$$\begin{aligned}\mathrm{cov}(f(\mathbf{u}_i), f(\mathbf{u}_j)) = N_1 \Bigg[& N_2 \exp\left\{ -\frac{1}{2\theta_1^2}(x_i - x_j)^2 \right\} \\ & \times \exp\left\{ -\frac{1}{2(1-\rho^2)} \left(\frac{1}{\theta_2^2}(\mathrm{M_{t}}_i - \mathrm{M_{t}}_j)^2 + \frac{1}{\theta_3^2}(\mathrm{JSF}_i - \mathrm{JSF}_j)^2 \right.\right. \\ & \left.\left. - \frac{2\rho}{\theta_2\theta_3}(\mathrm{M_{t}}_i - \mathrm{M_{t}}_j)(\mathrm{JSF}_i - \mathrm{JSF}_j) \right) \right\} + \sigma_i^2 \delta_{ij} \Bigg] .\end{aligned} \tag{6.8}$$

Here, N_1, N_2, θ_1, θ_2, θ_3, and ρ are the GP hyperparameters, σ_i is a noise parameter which accounts for the statistical uncertainty on the distribution bin underlying each training point, and δ_{ij} is the Kronecker delta function. The terms inside the exponentials specify the covariance between any two values of the shape as a function of their separation in x, $\mathrm{M_t}$, and JSF. The hyperparameters θ_1, θ_2, and θ_3 specify the lengthscales over which the GP shape is allowed to vary, and ρ is a correlation coefficient that couples the $\mathrm{M_t}$ and JSF parameters. The σ_i noise parameter incorporates the statistical uncertainty at each training point into the GP training process. The hyperparameters N_1 and N_2 control the normalization of the Gaussian and noise terms – N_1 specifies the overall normalization, and N_2 determines the relative balance between the two terms.

The values of the hyperparameters for the $\mathrm{M_{b\ell}}$, $\mathrm{M_{T2}^{bb}}$, and MAOS $\mathrm{M_{b\ell\nu}}$ distribution shapes are given in Table 6.1, with a discussion on hyperparameter estima-

tion provided in Section 6.3.1. The resulting GP shapes are shown in Chapter 7, where Figs. 7.2, 7.4, and 7.5 show the M_t-dependence and Fig. 7.6 shows the JSF-dependence of the shapes. These non-parametric shapes give rise to a likelihood fit with minimal bias, as shown by the MC validation studies in Figs. 7.7 and 7.8.

6.3.1 Additional details

Hyperparameter estimation

The values of the hyperparameters in Eq. (6.8) are determined with the help of a cross-validation likelihood fit [71], which is conducted for each observable separately. The lengthscale hyperparameters (θ_1, θ_2, and θ_3) must be small enough for the GP shape to pass through the training point values, and large enough for the shape to interpolate smoothly between them. Hyperparameters that are under-estimated satisfy the former criterion, but cause *overfitting* to occur in the resulting GP shape. This creates a noisy interpolation between training points, and may lead to bias in the measured value of M_t and its uncertainties. The GP shapes employed in Chapter 7 are checked for overfitting effects using several techniques, including one technique based on the local shape sensitivity function described in Chapter 8.

In the cross-validation method used to find optimal values for the hyperparameters, the set of training points is split into k disjoint sets – the first $(k-1)$ sets are used to train the GP shape, and the k^{th} set is used to validate the resulting shape. The k disjoint sets are chosen so that each contains points spread evenly across the total GP range. Given all training points $\mathbf{u}_0, \ldots, \mathbf{u}_N$, the i^{th} cross-validation set is populated with the points $\mathbf{u}_i, \mathbf{u}_{k+i}, \mathbf{u}_{2k+i}$, etc. Then, in the cross-validation

procedure, the training and validation sets are iterated k times. A new validation set is used in each iteration, so that after k iterations all points on the shape are validated.

In the k-fold cross-validation scheme described above, it is desirable to make k as large as possible, given the computational resources available. The most extreme case, when $k = N$, is known as the leave-one-out cross-validation (LOO-CV) scheme – here, the validation set consists of only one point, and N different GP trainings are required to carry out the cross-validation scheme (in our case, $N = 2625$).

The LOO-CV scheme is based on the log probability,

$$\log p(y_i|\mathbf{C}_{-i}, \mathbf{y}_{-i}, \boldsymbol{\theta}) = -\frac{1}{2}\log(\sigma_{\text{stat}})_i^2 - \frac{(y_i - \mu_i)^2}{2(\sigma_{\text{stat}})_i^2} - \frac{1}{2}\log 2\pi, \tag{6.9}$$

which evaluates the consistency of the shape with mean μ_i and statistical variance $(\sigma_{\text{stat}})_i^2$ at the validation point y_i. Here, the covariance $\mathbf{C}_{-i}$ and vector $\mathbf{y}_{-i}$ consist of all training points except the point y_i which is reserved for validation. Now that we have defined the log probability for a single iteration of the LOO-CV method, the total likelihood is given by

$$\mathcal{L}_{\text{LOO}}(\mathbf{C}_N, \mathbf{y}, \boldsymbol{\theta}) = \sum_{i=1}^{N} \log p(y_i|\mathbf{C}_{-i}, \mathbf{y}_{-i}, \boldsymbol{\theta}) \tag{6.10}$$

where we are, in effect, cycling through N different trainings of the GP shape.

The computational challenge here lies in computing the value of μ_i in Eq. (6.9) for each of the N trainings of the shape. This requires the inversion of matrix $\mathbf{C}_{-i}$ for each value of i, as evident by Eqs. (6.4) and (6.5) (effectively, the covariance matrix $\mathbf{C}_N$ with one row and column left out). To greatly simplify this task, the matrix algebra technique of inversion by partitioning can be implemented, so that

Table 6.1: GP hyperparameter values used to model the $M_{b\ell}$, M_{T2}^{bb}, and MAOS $M_{b\ell\nu}$ shapes.

	θ_1 (GeV)	θ_2 (GeV)	θ_3	ρ	N_1	N_2
$M_{b\ell}$	10.8	10.8	0.22	0.70	1.0×10^{-6}	1.14
M_{T2}^{bb}	8.1	8.1	0.14	0.69	1.0×10^{-6}	1.84
$M_{b\ell\nu}$	8.5	30	0.20	0.68	1.8×10^{-6}	0.42

the value of μ_i is expressed in terms of $\mathbf{C}_N$ instead of $\mathbf{C}_{-i}$. This gives the mean and variance

$$\mu_i = y_i - [\mathbf{C}_N^{-1}\mathbf{y}]_i / [\mathbf{C}_N^{-1}]_{ii} \tag{6.11}$$

$$\sigma_i^2 = 1/[\mathbf{C}_N^{-1}]_{ii}, \tag{6.12}$$

where it is clear that $\mathbf{C}_{-i}$ and $\mathbf{y}_{-i}$ no longer play a direct role. In Eq. (6.11), σ_i^2 is the variance of the GP shape, analogous to σ_{N+1}^2 in Eq. (6.5). This GP variance can be used in Eq. (6.9) to construct the CV likelihood. For the analysis in Chapter 7, the variance corresponding to the statistical uncertainty on each training point, σ_{stat}^2, is used instead.

The GP hyperparameter values chosen for the $M_{b\ell}$, M_{T2}^{bb}, and MAOS $M_{b\ell\nu}$ distribution shapes are given in Table 6.1. The lengthscale hyperparameters θ_1 and θ_2 have values of order 10 GeV. The hyperparameter θ_3 related to the JSF is smaller, of order 0.1, reflecting the fact that the JSF is a fractional quantity with variations of $\pm 3\%$ in the likelihood fit. The ρ hyperparameter is determined to be ~ 0.7, indicating a significant correlation between the M_t and JSF parameters. The hyperparameter N_1 is small due to the normalization of the distribution shapes to unity, and N_2 is of order unity to provide an appropriate balance between the Gaussian term and noise term in the GP kernel.

Shape normalization

The training points used to construct each GP shape are obtained from binned distributions that are normalized to unity. However, the normalization of the GP shape itself may deviate slightly from unity due to minor imperfections in shape modeling. To mitigate this effect, the GP shape normalization is recomputed using a numerical technique for each value of M_t and JSF at which the shape is evaluated. In a likelihood fit, the normalization is recomputed for every variation of the fit parameters.

Noise in null regions of distributions

In defining the range of a GP shape in the observable x, it is important to exclude regions where the corresponding distribution goes to zero. In these regions, the GP shape may fluctuate about zero, sometimes attaining negative values. Such negative values can be mitigated, for example by taking the absolute value of the GP shape. However, this approach results in discontinuities that may affect the minimization algorithms used in likelihood fitting. Such complications can be avoided simply by excluding null regions in the GP shape.

CHAPTER 7

MEASUREMENT OF THE TOP QUARK MASS IN THE DILEPTONIC DECAY CHANNEL

Hofstadter's Law: It always takes longer than you expect, even when you take into account Hofstadter's Law.

– Douglas Hofstadter

In this chapter, we report a measurement of the top quark mass in the dileptonic $t\bar{t}$ decay channel, $t\bar{t} \to (b\ell^+\nu)(\bar{b}\ell^-\bar{\nu})$, shown in Fig. 7.1. This measurement is documented in Ref. [73]. The dileptonic $t\bar{t}$ decay topology presents a unique challenge in mass measurement arising primarily from the presence of two neutrinos in the final state. While the missing $\vec{p}_{\mathrm{T}}$ vector of a single neutrino can be inferred from the missing transverse momentum, $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$, the allocation of $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ among two neutrinos is unknown a priori. For this reason, the dileptonic $t\bar{t}$ system is kinematically underconstrained, and mass determination cannot be conducted on an event-by-event basis. Instead, the mass of the parent top quarks in the dileptonic $t\bar{t}$ system must be extracted from kinematic distributions over an ensemble of events, with the help of appropriate observables and reconstruction techniques.

The measurement described here is based on a set of observables that have been proposed specifically for mass reconstruction in underconstrained decay topologies. These observables include the invariant mass, $M_{b\ell}$, of the $b\ell$ system, a 'stransverse mass' variable, M_{T2}^{bb}, constructed with the b and $\bar{b}$ daughters of the $t\bar{t}$ system [74–76], and the invariant mass of the $b\ell\nu$ system, $M_{b\ell\nu}$, where the neutrino momenta are estimated by the M_{T2}-assisted on-shell (MAOS) reconstruction technique [77].

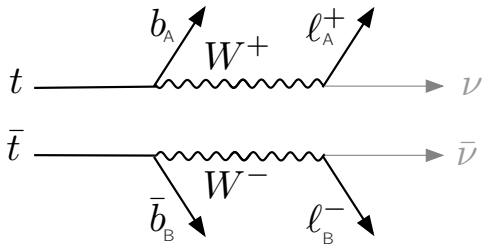

Figure 7.1: The dileptonic $t\bar{t}$ topology contains two identical branches, each terminating in one b jet, one electron or muon, and one neutrino.

Distributions of $M_{b\ell}$ and M_{T2} in dileptonic events contain a sharp edge descending to a kinematic endpoint, the location of which is sensitive to the value of M_t. In a previous measurement involving the author, masses of the top quark, W boson, and neutrino were extracted in a simultaneous fit using the endpoints of these distributions in dileptonic $t\bar{t}$ events [29]. The MAOS $M_{b\ell\nu}$ observable builds on M_{T2} by exploiting the neutrino momenta estimates that are a by-product of the M_{T2} algorithm. Its sensitivity to the top quark mass lies primarily in the edge to the left of its well-defined peak region, which is located approximately at the value of M_t. The $M_{b\ell}$, M_{T2}, and MAOS $M_{b\ell\nu}$ observables are described in more detail in Section 7.2.

One of the dominant sources of systematic error limiting the precision of this measurement comes from the overall uncertainty in jet energy scale (JES) determination. To address the JES uncertainty, we introduce a technique that uses the $M_{b\ell}$ and M_{T2} observables to determine an overall jet energy scale factor (JSF) simultaneously with the top quark mass extraction, where the JSF is defined as

a multiplicative factor scaling the four-vectors of all jets in the event. In parallel, there has been recent progress in developing approaches that take advantage of the $W \to jj$ decay in all-hadronic and semileptonic $t\bar{t}$ topologies to determine the JSF [28]. Because light-quark jets from the W boson decay are used to calibrate the energy scale of b jets stemming from the t and $\bar{t}$, these methods are sensitive to flavor-dependent uncertainties that emerge from differences in the response of b jets and light-quark jets. The method featured here determines the JSF in the dileptonic $t\bar{t}$ channel without relying on a hadronic W boson decay. Instead, it achieves sensitivity to the JSF through the kinematic differences between b jets, which are subject to JSF scaling, and leptons, which are not. Because it does not use light quarks from a hadronic W boson decay, this approach is robust against flavor-dependent JES uncertainties.

To model the $M_{b\ell}$, M_{T2}, and MAOS $M_{b\ell\nu}$ distribution shapes, we employ the GP regression technique described in Section 6.3. This technique is non-parametric, and thus largely model-independent. It is an effective method for modeling distribution shapes when no theoretical guidance is available to specify a functional form. The GP shapes can conveniently be modeled as functions of multiple variables. In this analysis, three variables are used – the value of the observable, M_t, and the JSF. The shapes are determined using simulated events generated with seven different values of M_t ranging from 166.5 GeV to 178.5 GeV, and with five values of JSF, ranging from 0.97 to 1.03, applied to the jets in each event. Each shape ultimately models the distributions of the observables together with their evolution in M_t and in JSF.

7.1 Datasets and event selection

We select dileptonic $\mathrm{t\bar{t}}$ events from a data set recorded during 2012, corresponding to an integrated luminosity of $19.7 \pm 0.5\,\mathrm{fb}^{-1}$ [78] and $\sqrt{s} = 8$ TeV. Events are required to pass one of several triggers that require at least two leptons, ee, eμ, or $\mu\mu$, where the leading (higher-p_T) lepton satisfies $p_\mathrm{T} > 17\,\mathrm{GeV}$ and the sub-leading lepton satisfies $p_\mathrm{T} > 8\,\mathrm{GeV}$.

The particle-flow (PF) algorithm [53–55] is used to reconstruct and identify each individual particle in an event by combining information from various elements of the CMS detector. Electron candidates are reconstructed by matching a cluster of energy deposits in the ECAL to a reconstructed track [56]. They are required to satisfy $p_\mathrm{T} > 20\,\mathrm{GeV}$ and $|\eta| < 2.5$. Muon candidates are reconstructed in a global fit that combines information from the silicon tracker and muon system [58], and must have $p_\mathrm{T} > 20\,\mathrm{GeV}$ and $|\eta| < 2.4$. A requirement on the relative isolation is imposed inside a cone $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ around each lepton candidate, where a parameter $I_\mathrm{rel} = \sum p_{\mathrm{T}\,i}/p_\mathrm{T}^{\ell}$ is defined. Here, the sum includes all reconstructed PF candidates inside the cone (excluding the lepton itself), and p_T^{ℓ} is the lepton p_T. Electron candidates are required to have $I_\mathrm{rel} < 0.15$ with $\Delta R < 0.3$, and muon candidates must satisfy $I_\mathrm{rel} < 0.2$ with $\Delta R < 0.4$. Events selected offline are required to contain exactly two such leptons, ee, eμ, or $\mu\mu$ with opposite charge. For events containing an $\mathrm{e^+e^-}$ or $\mu^+\mu^-$ pair, contributions from Z-boson decays are suppressed by requiring that $|\mathrm{M_Z} - \mathrm{M}_{\ell\ell}| > 15\,\mathrm{GeV}$, and contributions from low-mass resonances are suppressed by requiring that $\mathrm{M}_{\ell\ell} > 20\,\mathrm{GeV}$.

Hadronic jets are clustered from PF candidates with the infrared and collinear safe anti-k_t algorithm [59], with a size parameter ΔR of 0.5, as implemented in the FastJet package [60]. The jet momentum is determined as the vectorial sum of

all particle momenta in this jet. Corrections to the JES and jet energy resolution (JER) are derived using MC simulation, and are confirmed with measurements of the energy balance in quantum chromodynamics (QCD) dijet, photon+jet, and Z+jet events [79]. Muons and electrons, as well as charged hadrons originating from pileup interactions are not included in the jet reconstruction. Jets originating from the hadronization of b quarks are identified with a combined secondary vertex (CSV) b tagging algorithm [25], combining information from the jet's secondary vertex with the impact parameter significances of its constituent tracks. Jets selected by the algorithm are referred to as b jets. Events selected offline are required to contain at least two b jets satisfying $p_{\mathrm{T}} > 30\,\mathrm{GeV}$ and $|\eta| < 2.5$.

The missing transverse momentum vector is defined as $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}} = -\sum \vec{p}_{\mathrm{T}i}$, where the sum includes all reconstructed PF candidates in an event. Its magnitude is referred to as $E_{\mathrm{T}}^{\mathrm{miss}}$. Corrections to the JES and JER are propagated into the $E_{\mathrm{T}}^{\mathrm{miss}}$, as well as an offset correction that accounts for pileup interactions. An additional correction mitigates a mild azimuthal dependence that is observed in the reconstructed $E_{\mathrm{T}}^{\mathrm{miss}}$. To suppress contributions from Z boson decays, events containing an $\mathrm{e^+e^-}$ or $\mu^+\mu^-$ pair are required to contain $E_{\mathrm{T}}^{\mathrm{miss}} > 40\,\mathrm{GeV}$.

Simulated $\mathrm{t\bar{t}}$ signal events are generated with the MadGraph 5.1.5.11 matrix element generator [66], combined with MadSpin to include spin correlations of the top quark decay products [80], pythia 6.426 with the Z2* tune for parton showering [64], and tauola for the decay of τ leptons [81]. Parton distribution functions are described by the CTEQ6L1 set [82]. The $\mathrm{t\bar{t}}$ signal events are generated with seven different values of $\mathrm{M_t}$ ranging from 166.5 to 178.5 GeV. The contribution from the tW single-top process is simulated with powheg 1.380 [83–86], where the value of $\mathrm{M_t}$ is assumed to be 172.5 GeV. Background events from W+jets

and Z+jets production are generated with MadGraph 5.1.3.30, and contributions from WW, WZ, ZZ processes are simulated with pythia 6.426 using the Z2* tune. The CMS detector response to the simulated events is modelled with Geant4 [69]. All background processes are normalized to their predicted cross sections [87–91].

With the requirements outlined above, 41 640 $\mathrm{t\bar{t}}$ candidate events are selected in data. The sample composition is estimated in simulation to be 95% $\mathrm{t\bar{t}}$, 4% single top quark, and 1% other processes including diboson, W+jets, and Drell Yan.

7.2 Observables

The observables featured in this study have been developed for physics scenarios where undetected particles, such as neutrinos, carry away a portion of the kinematic information necessary for full event reconstruction. In the dileptonic $\mathrm{t\bar{t}}$ system, distributions in these observables contain endpoints, edges, and peaks that are sensitive to the top quark mass. The observables are described in more detail below.

7.2.1 $\mathrm{M_{b\ell}}$

The $\mathrm{M_{b\ell}}$ observable is defined by:

$$\mathrm{M^2_{b\ell}} = (p_\mathrm{b} + p_\ell)^2, \tag{7.1}$$

where p_b and p_ℓ are four vectors corresponding to a b jet and lepton, respectively. The bℓ pairs underlying each value of $\mathrm{M_{b\ell}}$ are chosen out of four possible combina-

tions by an algorithm described below. The $M_{b\ell}$ observable contains a kinematic endpoint that occurs when the b jet and lepton are directly back-to-back in the top rest frame. The location of this endpoint, $(M_{b\ell})_{max}$, is a function of the masses involved in the decay:

$$(M_{b\ell})_{max} = \frac{\sqrt{(M_t^2 - M_W^2)(M_W^2 - M_\nu^2)}}{M_W}. \quad (7.2)$$

With $M_t = 172.5$ GeV, $M_W = 80.4$ GeV, and $M_\nu = 0$ GeV, we have $(M_{b\ell})_{max} = 152.6\,\text{GeV}$. Although this endpoint is a theoretical maximum on the value of $M_{b\ell}$ at leading order, events are still observed beyond this value due to higher-order corrections, resolution effects, finite particle widths, and background contamination.

The $M_{b\ell}$ distribution is shown in data and MC simulation in Fig. 7.2 (left), with a breakdown of signal and background events shown in the simulation. The 'signal' category includes $t\bar{t}$ dilepton decays where both b jets are correctly identified by the b tagging algorithm. The background categories include: 'mistag' dilepton decays where a light quark or gluon jet is incorrectly selected by the b tagging algorithm; 'tau decays' where dilepton events include at least one τ in the final state subsequently decaying leptonically; and 'hadronic decays' that include events where at least one of the top quarks decays hadronically. The 'non-ttbar bkg' category includes single top quark, diboson, W+jets, and Drell Yan processes.

The sensitivity of the $M_{b\ell}$ observable to the value of M_t is demonstrated in Fig. 7.2 (right), where $M_{b\ell}$ shapes corresponding to three values of the top quark mass in MC simulation (M_t^{MC}) are shown. The variation between these shapes reveals regions of the $M_{b\ell}$ distribution that are sensitive to the value of M_t, such as the edges to the left and right of the $M_{b\ell}$ peak, and regions that are not sensitive, such as the stationary point where the three shapes intersect. To provide

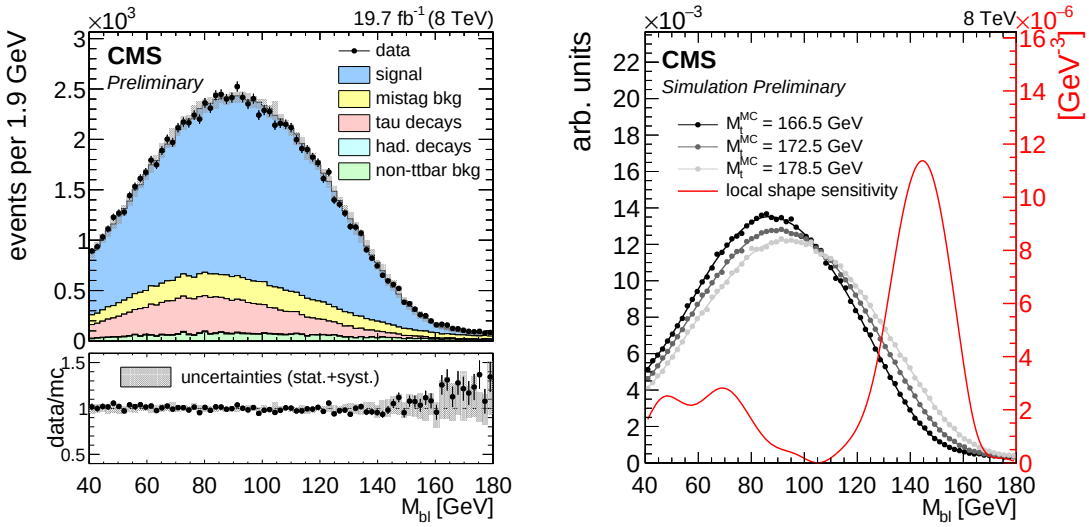

Figure 7.2: (Left) the $M_{b\ell}$ distribution in data and simulation with $M_t^{MC} = 172.5\,\text{GeV}$, normalized to the number of events in the 8 TeV data set corresponding to an integrated luminosity of $19.7 \pm 0.5\,\text{fb}^{-1}$. Statistical and systematic uncertainties on the distribution in simulation are represented by the grey shaded area. A description of the systematic uncertainties is given in Section 7.5. (Right) the $M_{b\ell}$ distribution shapes in simulation corresponding to three values of M_t^{MC} are shown in grey. The 'local shape sensitivity' function, described in Chapter 8, is shown in red.

a quantitative description of these effects, we introduce a 'local shape sensitivity' function, also known as the Fisher information density, represented by the red line in Figs. 7.2, 7.4, and 7.5. This function conveys the sensitivity of an observable at a specific point on its shape. For the $M_{b\ell}$ observable, the local shape sensitivity function peaks near the kinematic endpoint ($M_{b\ell} \sim 150\,\text{GeV}$), and has a zero value at the stationary point ($M_{b\ell} \sim 105\,\text{GeV}$). The integral of this function over its range is proportional to $1/\sigma^2_{M_t}$, where σ_{M_t} is the statistical uncertainty on a measurement of M_t. A full description of the local shape sensitivity function is given in Chapter 8.

b jet, lepton combinatorics

The two b jets and two leptons stemming from each $t\bar{t}$ decay give rise to a two-fold matching ambiguity, with two correct and two incorrect $b\ell$ pairings possible in each event. Pairings in which the b jet and lepton emerge from different top quarks do not necessarily obey the upper bound described in Eq. (7.2), and thus do not have a clean kinematic endpoint in $M_{b\ell}$. Although a priori it is impossible to distinguish between correct and incorrect pairings, one possible approach is to select the smallest two $M_{b\ell}$ values in each event. This way, the kinematic endpoint of the distribution is preserved – even if the smallest two values do not correspond to the correct pairings, they are guaranteed to fall below the correct pairings, which do respect the endpoint. In this analysis, we employ a slightly more sophisticated matching technique, introduced in Ref. [29], where either two or three $b\ell$ pairs are selected in each event.

By selecting either two or three $b\ell$ pairs in each event, the technique employed in this analysis has the benefit of increased statistics, while preserving the kinematic endpoint of $M_{b\ell}$. Although they are not necessarily the correct pairs, the corresponding $M_{b\ell}$ values are guaranteed by construction to be less than or equal to those of the correct pairs. The matching technique is based on the following prescription:

1. Match each b jet with the lepton that produces the lower $M_{b\ell}$ value.
2. Match each lepton with the b jet that produces the lower $M_{b\ell}$ value.

This recipe produces either two or three values of $M_{b\ell}$. In the latter case, two different leptons may be successfully paired with the same b jet, and vice versa. Note the difference between this approach and the simpler approach of choosing

the smallest two values of $M_{b\ell}$. The smallest two values of $M_{b\ell}$ do not necessarily incorporate both b jets and both leptons in the event. For example, this could occur if both b jets are matched to a single lepton. In these cases, the next largest $M_{b\ell}$ value is also needed to ensure both b jets and both leptons from the event are used. This approach for bℓ-matching provides the statistical benefit of two or three pairings per event, while preserving the kinematic endpoint that is central to the $M_{b\ell}$ observable.

7.2.2 M_{T2}

The M_{T2} observable [74, 75] is based on the transverse mass, M_T. The transverse mass of the W boson in a $W \to \ell\nu$ decay is given by

$$M_T = \sqrt{m_\ell^2 + m_\nu^2 + 2(E_{T\ell}E_{T\nu} - \vec{p}_{T\ell} \cdot \vec{p}_{T\nu})}\,, \tag{7.3}$$

where $E_{Tx}^2 = m_x^2 + \vec{p}_{Tx}^2$ for $x \in \{\ell, \nu\}$. This quantity exhibits a kinematic endpoint at the parent mass, M_W, which occurs in configurations when both the lepton and neutrino momenta lie entirely in the transverse plane (up to a common longitudinal boost).

The dileptonic $t\bar{t}$ system has an additional layer of decays, with $t \to Wb$ in the first step followed by $W \to \ell\nu$ in the second. The result is an event topology with two identical branches, $t \to b\ell^+\nu$ and $\bar{t} \to \bar{b}\ell^-\bar{\nu}$, each with a visible ($b\ell$) and invisible ($\nu$) component. In this case, one value of M_T can be computed for each branch. The invisible particle momentum associated with each branch, however, is not known. While for a semileptonic $t\bar{t}$ decay, with only one $W \to \ell\nu$ decay, the neutrino p_T is estimated from the E_T^{miss} in the event, a dileptonic $t\bar{t}$ decay includes two neutrinos, for which the allocation of E_T^{miss} between them is unknown.

The M_{T2} observable is an extension of M_T for a system with two identical decay branches, such as those in the dileptonic $t\bar{t}$ system. Here, the invisible particle momenta, $\vec{p}_T^{\,a}$ and $\vec{p}_T^{\,b}$, must add up to the total $\vec{p}_T^{\,miss}$. The strategy of M_{T2} is to impose this constraint on the invisible particle momenta, while also performing a minimization in order to preserve the kinematic endpoint of M_T. For a general event with a symmetric decay topology, M_{T2} is defined as:

$$M_{T2} = \min_{\vec{p}_T^{\,a}+\vec{p}_T^{\,b}=\vec{p}_T^{\,miss}} [\max\{M_T^a, M_T^b\}], \tag{7.4}$$

where M_T^a and M_T^b correspond to the two decay branches. If the invisible particle mass is known it can be incorporated into the M_{T2} calculation as well, yielding an endpoint at the parent particle mass. Although the final values of $\vec{p}_T^{\,a}$ and $\vec{p}_T^{\,b}$ are typically treated as intermediate quantities in the M_{T2} algorithm, they are employed as neutrino $\vec{p}_T$ estimates in the MAOS reconstruction technique described in Section 7.2.3.

M_{T2} Subsystems

In the $t\bar{t}$ system, there are several ways in which M_{T2} can be computed, depending on how the decay products are grouped together. The M_{T2} observable classifies them into three categories: upstream, visible, and child particles [92]. The child particles are those at the end of the decay chain that are unobservable or simply treated as unobservable. In the latter case, the child particle momenta are added to the $\vec{p}_T^{\,miss}$ vector. The visible particles are those whose transverse momenta are measured and used in the calculations; and the upstream particles are those from further up the decay chain, including any initial state radiation (ISR) accompanying the hard collision.

In general, the child, visible, and upstream particles may actually be collections

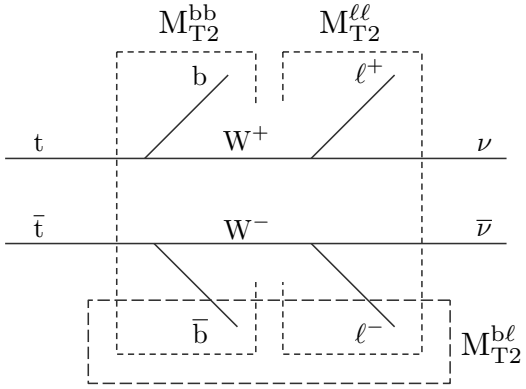

Figure 7.3: M_{T2} subsystems in the dileptonic $t\bar{t}$ event topology.

of objects, creating three possible subsystems in the dileptonic $t\bar{t}$ event topology. For simplicity, we refer to the corresponding M_{T2} observables as M_{T2}^{bb}, $M_{T2}^{\ell\ell}$, and $M_{T2}^{b\ell}$:[1]

- The $\mathbf{M_{T2}^{\ell\ell}}$ **observable** uses the two leptons as visible particles, treating the neutrinos as invisible child particles (which they are), and combining the b jets with all other upstream particles in the event.
- The $\mathbf{M_{T2}^{bb}}$ **observable** uses the b jets as visible particles, and treats the W bosons as child particles (ignoring the fact that their charged daughter leptons are indeed observable). It considers only ISR jets as generators of upstream momentum.
- The $\mathbf{M_{T2}^{b\ell}}$ **observable** combines the b jet and the lepton to form a single visible system, and takes the neutrinos as the invisible particles.

The subsystem observable M_{T2}^{bb} is employed in this study to complement the observable $M_{b\ell}$. The M_{T2}^{bb} observable contains an endpoint at the value of M_t, and

[1]These observables are identical, respectively, to $M_{T2}^{(2,2,1)}$, $M_{T2}^{(2,1,0)}$, $M_{T2}^{(2,2,0)}$ of Ref. [92], and μ_{bb}, $\mu_{\ell\ell}$, $\mu_{b\ell}$ of Ref. [29].

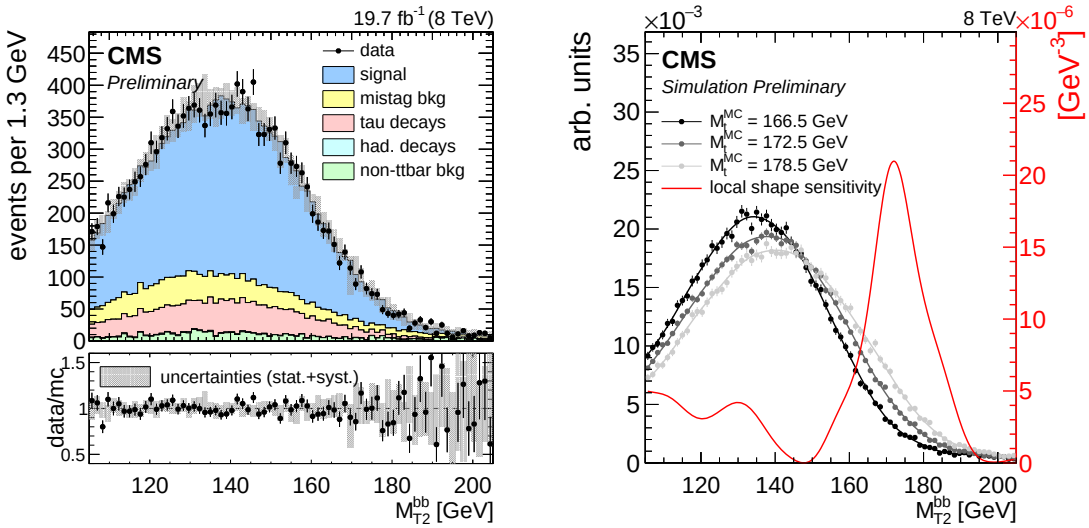

Figure 7.4: (Left) the M_{T2}^{bb} distribution in data and simulation with $M_t^{MC} = 172.5\,\text{GeV}$, normalized to the number of events in the 8 TeV data set corresponding to an integrated luminosity of $19.7 \pm 0.5\,\text{fb}^{-1}$. Statistical and systematic uncertainties on the distribution in simulation are represented by the grey shaded area. (Right) the M_{T2}^{bb} distribution shapes in simulation corresponding to three values of M_t^{MC} are shown in grey. The 'local shape sensitivity' function is shown in red.

can be combined with $M_{b\ell}$ to mitigate uncertainties due to the JES. This feature is discussed further in Section 7.3. The distribution of M_{T2}^{bb} and its sensitivity to the value of M_t are shown in Fig. 7.4. Although $M_{T2}^{\ell\ell}$ is not directly sensitive to M_t, the neutrino $\vec{p}_T$ estimates that are a by-product of its computation are used as an input into the MAOS $M_{b\ell\nu}$ reconstruction technique described below.

The M_{T2}^{bb} distribution employed in this analysis includes a kinematic requirement on the direction of the upstream $\vec{p}_T$, which must lie outside the opening angle between the two b jet $\vec{p}_T$ vectors in the event. This requirement primarily impacts events at low values of M_{T2}^{bb}, and its effect on the observable's sensitivity is small.

7.2.3 MAOS $\mathrm{M_{b\ell\nu}}$ reconstruction

The MAOS reconstruction technique employed in this analysis is based on the subsystem observable $\mathrm{M_{T2}^{\ell\ell}}$. In the $\mathrm{M_{T2}^{\ell\ell}}$ algorithm, an $\mathrm{M_T}$ variable (Eq. (7.3)) is constructed from the $\ell^+\nu$ and $\ell^-\overline{\nu}$ pairs corresponding to each of the $\mathrm{t\bar{t}}$ decay branches. Because the values of neutrino $\vec{p}_{\mathrm{T}}$ are unknown, a minimization is conducted over possible values consistent with the $\vec{p}_{\mathrm{T}}^{\,\mathrm{miss}}$ in each event (Eq. (7.4)).

The MAOS technique employs the neutrino $\vec{p}_{\mathrm{T}}$ values that are determined by the $\mathrm{M_{T2}^{\ell\ell}}$ minimization, but substitutes the $\mathrm{M_T}$ variable corresponding to each $\ell\nu$ pair with a full $\mathrm{b}\ell\nu$ invariant mass. Given the neutrino $\vec{p}_{\mathrm{T}}$ values, the remaining z-components of the momenta are obtained by enforcing the W mass on-shell requirement:

$$\mathrm{M}(\ell^+\nu) = \mathrm{M}(\ell^-\overline{\nu}) = \mathrm{M_W} = 80.4\,\mathrm{GeV}. \tag{7.5}$$

This yields a longitudinal momentum for each neutrino given by

$$p_{z\nu} = \frac{1}{E_{T\ell}^2}\left[p_{z\ell}A \pm \sqrt{p_{z\ell}^2 + E_{T\ell}^2}\sqrt{A^2 - (E_{T\ell}E_{T\nu})^2}\right], \tag{7.6}$$

where $A = \frac{1}{2}(\mathrm{M_W^2} + \mathrm{M_\nu^2} + \mathrm{M_\ell^2}) + \vec{p}_{\mathrm{T}\ell} \cdot \vec{p}_{\mathrm{T}\nu}$ [77]. Given these estimates for the neutrino three-momenta together with $\mathrm{M_\nu} = 0$, we have the required four vectors to construct an $\mathrm{M_{b\ell\nu}}$ invariant mass corresponding to the decay products of each top quark.

The quadratic equations underlying the W mass on-shell requirement provide two solutions for each value of $p_{z\nu}$, yielding a two-fold ambiguity for each neutrino momentum (Eq. (7.6)). In addition, there is a two-fold ambiguity resulting from the matching of b jets to $\ell\nu$ pairs in the construction of $\mathrm{b}\ell\nu$ invariant masses. (No matching ambiguity exists between leptons and neutrinos, since the $\ell^+\nu$ and $\ell^-\overline{\nu}$ pairs have been fixed by the $\mathrm{M_{T2}^{\ell\ell}}$ algorithm.) The combined four-fold ambiguity,

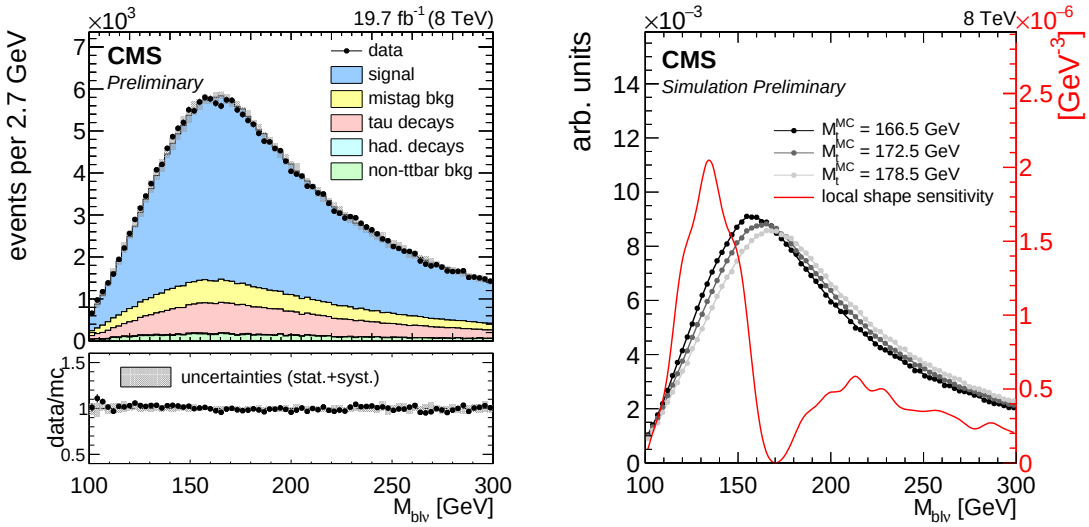

Figure 7.5: (Left) the MAOS $M_{b\ell\nu}$ distribution in data and simulation with $M_t^{MC} = 172.5\,\text{GeV}$, normalized to the number of events in the 8 TeV data set corresponding to an integrated luminosity of $19.7 \pm 0.5\,\text{fb}^{-1}$. Statistical and systematic uncertainties on the distribution in simulation are represented by the grey shaded area. (Right) the MAOS $M_{b\ell\nu}$ distribution shapes in simulation corresponding to three values of M_t^{MC} are shown in grey. The 'local shape sensitivity' function is shown in red.

along with two top quark decays in each event, gives eight possible values of $M_{b\ell\nu}$. In the measurement, all eight of the available values are used: for each $\ell\nu$ pair, this includes two neutrino $p_{z\nu}$ solutions, and two b-$\ell\nu$ matches. The distribution of MAOS $M_{b\ell\nu}$ and its sensitivity to the value of M_t is shown in Fig. 7.5.

7.3 Simultaneous determination of M_t and JSF

To mitigate the impact of JES uncertainties on the precision of this measurement, we introduce a technique that allows a JSF parameter to be fit simultaneously with the determination of M_t. The JSF is a constant multiplicative factor that calibrates the overall energy scale of reconstructed jets. Although it does not

account for the p_{T}- and η-dependent variations in jet response and pile up effects that are addressed by the JES, the overall energy scale captured in the JSF is the dominant component of uncertainty in the JES calibration.

The challenge in determining the JSF simultaneously with $\mathrm{M_t}$ stems from the large degree of correlation between these parameters. In the top quark decay, $\mathrm{t} \to \mathrm{b}\ell\nu$, the JSF directly affects the momentum of the b jet, and indirectly, the inferred momentum of the neutrino, by scaling all jets entering the $E_{\mathrm{T}}^{\mathrm{miss}}$ sum. The $\mathrm{M_t}$ parameter affects the momenta of these two particles in addition to the lepton produced in the top quark decay. In the context of observables and distribution shapes, variations in the $\mathrm{M_t}$ and JSF parameters cause shape changes that are difficult to distinguish. For this reason, a shape-based analysis using a single observable can be implemented to determine either $\mathrm{M_t}$ or JSF, but not both simultaneously.

To determine the $\mathrm{M_t}$ and JSF parameters simultaneously, we construct a likelihood that contains two distribution shapes corresponding to the $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ observables. In this configuration, variations in the parameters produce shifts in each individual shape, but they also create a relative shift between the shapes that provides the additional constraint needed for a simultaneous fit of $\mathrm{M_t}$ and JSF. The dependence of the $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ distribution shapes on $\mathrm{M_t}$ is shown in Figs. 7.2 and 7.4, and their dependence on the JSF is shown in Fig. 7.6. The difference in response between the $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ shapes to the JSF parameter is rooted in the reconstructed objects underlying the $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ observables – while each value of $\mathrm{M_{b\ell}}$ uses one b jet and one lepton, each value of $\mathrm{M_{T2}^{bb}}$ uses two b jets and no leptons for the visible system, thus exhibiting a stronger dependence on the JSF. The event-by-event likelihood fit used in this measurement is described in

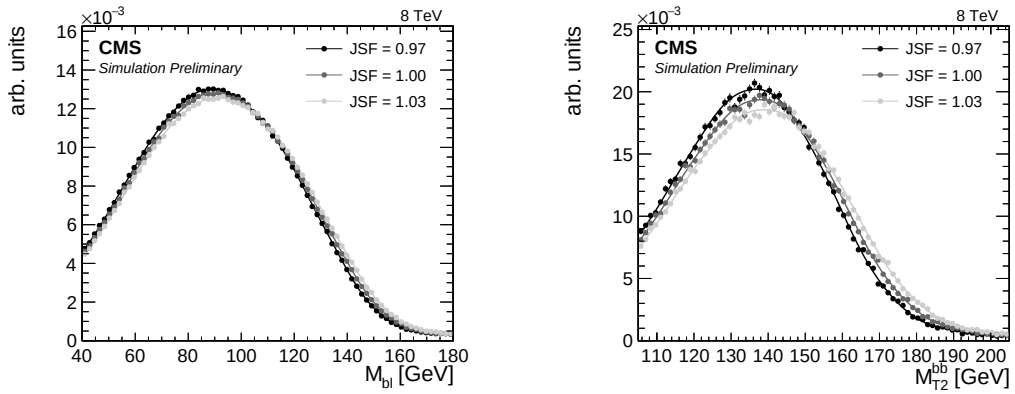

Figure 7.6: The (left) $M_{b\ell}$ and (right) M_{T2}^{bb} distributions in MC with M_t = 172.5 GeV for several values of JSF.

more detail in Section 7.4.

7.4 Fit strategy

This measurement employs an event-by-event maximum likelihood fit using the $M_{b\ell}$, M_{T2}^{bb}, and MAOS $M_{b\ell\nu}$ observables described in Section 7.2, along with the GP shape estimate technique described in Section 6.3. The MC datasets used to train the GP shapes include the $t\bar{t}$ signal and background processes described in Section 7.1.

The likelihood constructed from a single observable, x, is given by:

$$\mathcal{L}^x(M_t, \text{JSF}) = \prod_i f(x_i|M_t, \text{JSF}). \tag{7.7}$$

Here, the distribution shape f depends on the value of the free parameters M_t and JSF, and expresses the likelihood of drawing some event i where the value of the observable is x_i. It is normalized to unity over its range for all values of M_t and JSF. The parameters M_t and JSF are varied in the fit to maximize the value of the likelihood.

A likelihood containing two observables, x_1 and x_2, is constructed as a product of individual likelihoods:

$$\begin{aligned}\mathcal{L}(\mathrm{M_t}, \mathrm{JSF}) &= \mathcal{L}^{x_1}(\mathrm{M_t}, \mathrm{JSF}) \cdot \mathcal{L}^{x_2}(\mathrm{M_t}, \mathrm{JSF}) \\ &= \prod_i f(x_{1i}|\mathrm{M_t}, \mathrm{JSF}) \cdot f(x_{2i}|\mathrm{M_t}, \mathrm{JSF})\end{aligned} \tag{7.8}$$

A pseudo-experiment technique, described below, is employed to mitigate the effects of correlation between the observables x_1 and x_2. This analysis employs three different versions of the likelihood fit:

1. The **1D fit** uses the $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ observables to determine $\mathrm{M_t}$; JSF is constrained to be unity.
2. The **2D fit** also uses $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ but imposes no constraints and determines $\mathrm{M_t}$ and JSF simultaneously.
3. The **MAOS fit** uses the $\mathrm{M_{T2}^{bb}}$ and $\mathrm{M_{b\ell\nu}}$ observables to determine $\mathrm{M_t}$; JSF is constrained to be unity.

Among these versions, the 1D fit provides the best overall precision on the value of $\mathrm{M_t}$. The 2D fit mitigates the JES uncertainties, which are the largest source of systematic error in the 1D approach. The MAOS fit is expected to yield results similar to the 1D fit, and is presented as a viable alternative that substitutes the $\mathrm{M_{b\ell}}$ observable for MAOS $\mathrm{M_{b\ell\nu}}$. The best overall precision on $\mathrm{M_t}$ is given by a combination of the 1D and 2D fits, which is discussed in Section 7.4.1. The fit results are discussed in Section 7.6.

The central value and statistical uncertainty on $\mathrm{M_t}$ and JSF are determined using the bootstrapping technique [93]. This method is based on pseudo-experiments rather than the shape of the total likelihood (Eq. (7.8)) near its maximum, and

thus mitigates the effects of correlation between the two observables, x_1 and x_2, in the likelihood. The pseudo-experiments are constructed by resampling the full dataset with replacement, where the size of each pseudo-experiment is fixed to have the number of events in data (41 640 events). In simulation, the probability of selecting a particular event is proportional to its weight, containing the relevant cross sections, as well as corrections for MC modeling and object reconstruction efficiencies. Because the resampling is conducted with replacement, a single event may be selected more than once for any given pseudo-experiment.

The performance of the likelihood fitting approach described above is evaluated using events in simulation, where the true values of M_t and JSF are known. The fit is conducted using seven different values of M_t^{MC} ranging from 166.5 to 178.5 GeV for each version of the likelihood fit. The results of this performance study are shown in Figs. 7.7 and 7.8. It is evident that the likelihood fits are consistent with zero bias, showing that the GP shape modeling technique accurately captures the distribution shapes and their evolution over several values of M_t^{MC}. For this reason, no calibration of the fit is necessary for an unbiased determination of the M_t and JSF parameters.

7.4.1 Combination of 1D and 2D fits

The 1D and 2D fits discussed above have differing sensitivities to various sources of systematic uncertainty in this measurement. Although the 2D fit successfuly mitigates the JES uncertainties, which dominate in the 1D fit, other uncertainties in the 2D method are larger and cause the total precision to worsen. The best overall precision on the value of M_t is provided by a **hybrid fit**, defined as a linear combination of the 1D and 2D fits. The measured value of M_t in the hybrid fit is

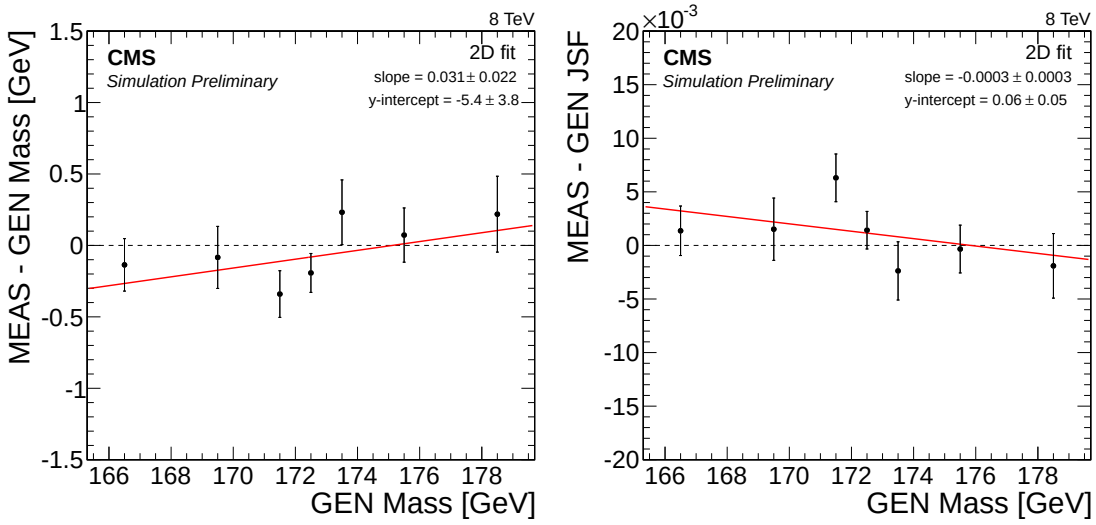

Figure 7.7: Likelihood fit results using 50 pseudo-experiments in MC simulation, with values of M_t^{MC} ranging from 166.5 to 178.5 GeV. A calibration curve of the form $y = ax + b$ is determined for each fit configuration. Measured values of (left) M_t and (right) JSF are shown for the 2D fit.

given by:

$$M_t^{hyb} = w_{hyb} M_t^{1D} + (1 - w_{hyb}) M_t^{2D}, \tag{7.9}$$

where the parameter w_{hyb} determines the relative weight between the 1D and 2D fits in the combination. The value of M_t^{hyb} and its statistical uncertainty are extracted using bootstrap pseudo-experiments, as described in Section 7.4. In each pseudo-experiment, the measured value of M_t^{hyb} is given by a linear combination of the measured M_t^{1D} and M_t^{2D} values (Eq. (7.9)). A value of $w_{hyb} = 0.8$ is found to achieve the best total precision on M_t. The performance of the hybrid fit, evaluated using MC samples corresponding to seven values of M_t^{MC}, is shown in Figs. 7.7 and 7.8.

As an independent cross-check of the linear combination method described above, an alternate hybrid fit, introduced in Ref. [28], was constructed by adding a Gaussisan constraint, $P(\text{JSF})$, to the likelihood defined in Eq. (7.8). The hybrid

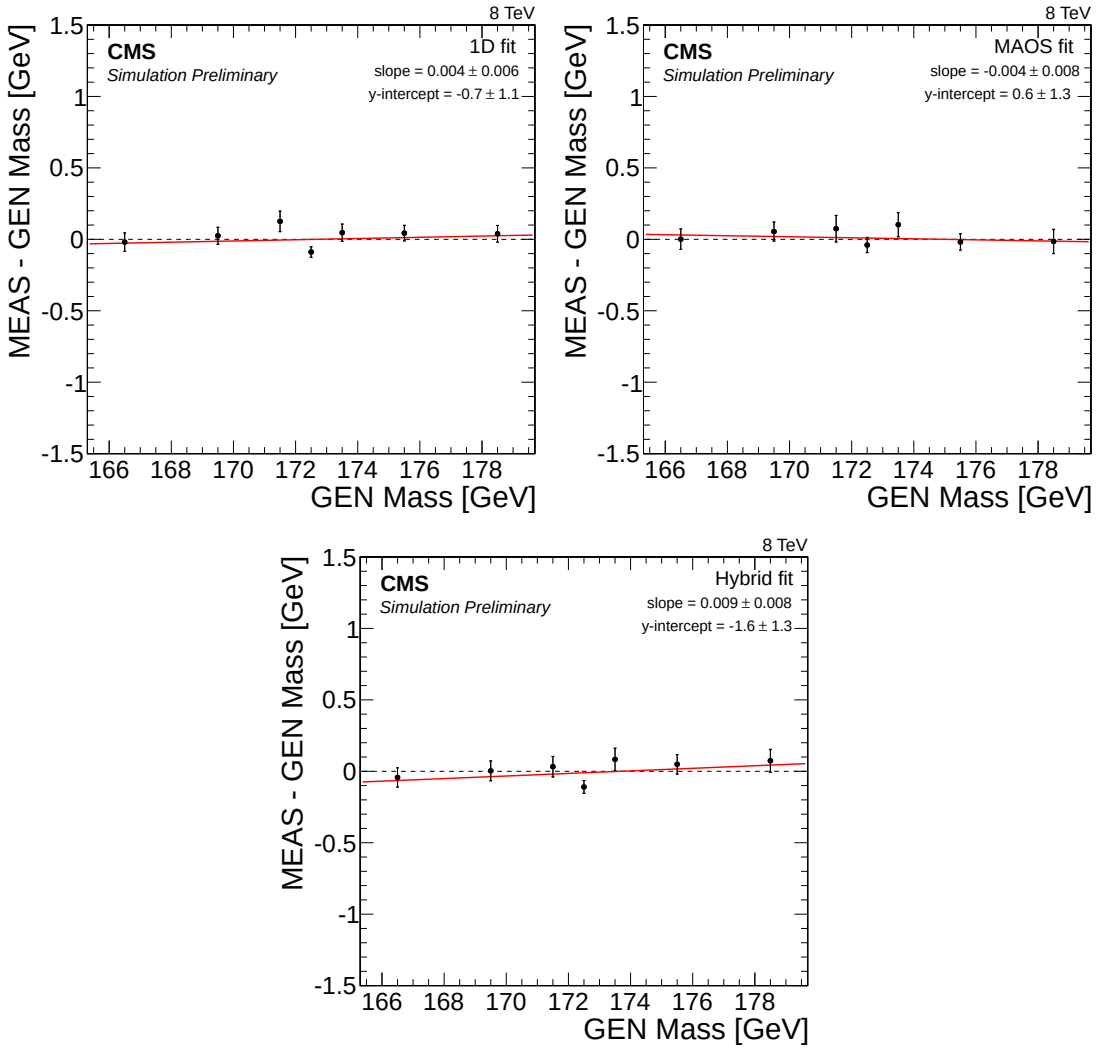

Figure 7.8: Likelihood fit results using 50 pseudo-experiments in MC simulation, with values of M_t^{MC} ranging from 166.5 to 178.5 GeV. A calibration curve of the form $y = ax + b$ is determined for each fit configuration. Measured values of M_t are shown for the (top left) 1D, (top right) MAOS, and (bottom) hybrid fits.

likelihood is then given by $\mathcal{L}^{\rm hyb}(\mathrm{M_t}, \mathrm{JSF}) = P(\mathrm{JSF})\mathcal{L}(\mathrm{M_t}, \mathrm{JSF})$. The Gaussian $P(\mathrm{JSF})$ is has a mean at JSF = 1.0 and a width that is tuned to provide the desired value of $\mathrm{w_{hyb}}$. The results of this alternate approach are found to be consistent with the linear combination method discussed above.

7.5 Systematic uncertainties

The systematic uncertainties evaluated in this measurement are given in Table 7.1. The uncertainties include experimental effects from detector calibration and object reconstruction, and modeling effects mostly arising from the simulation of QCD processes in MC simulation. All uncertainties are determined by conducting the likelihood fit using events from MC simulation with the relevant parameters varied by $\pm 1\Delta$, where Δ is the uncertainty on a particular parameter. The difference in the measured top quark mass ($\delta \mathrm{M_t}$) or JSF (δJSF) is taken to be the corresponding systematic uncertainty. For uncertainties that are evaluated by comparing two or more independent MC samples, the values of $\delta \mathrm{M_t}$ and δJSF may be subject to statistical fluctuations. For this reason, if the value of $\delta \mathrm{M_t}$ or δJSF is smaller than its statistical uncertainty in a particular systematic variation, the latter is quoted as the systematic uncertainty. Finally, if a systematic uncertainty is one-sided, where both $+\Delta$ and $-\Delta$ variations produce $\delta \mathrm{M_t}$ or δJSF shifts of the same sign, the larger shift is symmetrized to obtain the corresponding systematic uncertainty.

In the hybrid fit, the systematic uncertainties are evaluated according to the linear combination in Eq. (7.9). In this approach, we assume that the uncertainty sources are fully correlated between the 1D and 2D fits, both of which use the same observables and an equivalent set of events. For each systematic variation,

this gives $\delta M_t^{hyb} = w_{hyb}\delta M_t^{1D} + (1 - w_{hyb})\delta M_t^{2D}$. This approach provides the smallest overall uncertainty, with the largest contributions stemming from the JES, b quark fragmentation modeling, and hard scattering scale. The next most precise result is given by the 1D fit, also dominated by the same sources of uncertainty. The JES uncertainties are successfully mitigated in the 2D fit. The 2D fit, however, is more sensitive to the uncertainties in the top quark p_T spectrum, matching scale, and underlying event tune, so the total systematic uncertainty for the 2D fit is larger than that of the 1D fit. The MAOS fit has a larger total systematic uncertainty than the 1D fit due to its sensitivity to the JES, top quark p_T spectrum, and b quark fragmentation modeling uncertainties. Further details on each source of systematic uncertainty are given below.

7.5.1 Jet energy scale

The JES uncertainty is evaluated separately for four components, which are then added in quadrature [94]. The 'Intercalibration' group arises from the modeling of radiation in the p_T- and η-dependent JES determination. The 'MPF in situ' category includes uncertainties stemming from the determination of the absolute JES using γ/Z+jet events. The 'Uncorrelated' component includes uncertainties due to detector effects and pileup. Finally, the 'Flavor' uncertainty stems from difference in the energy response between different jet flavors – it is a linear sum of contributions from the light quark, charm quark, bottom quark, and gluon response, which are estimated by comparing the Lund string fragmentation in PYTHIA 6 [64] and cluster fragmentation in HERWIG ++ [95] for each type of jet. All JES uncertainties are propagated into the reconstructed E_T^{miss} in each event.

7.5.2 Jet energy resolution

The resolution of jets is known to be underestimated in MC simulation compared to data. This effect is corrected with a set of scale factors that are used to smear the jet four-vectors to broaden their resolutions. The scale factors are determined in bins of η. Here, they are varied within their uncertainties, which are typically 2.5-5% of the scale factor. The effect of these variations is also propagated into the $E_{\mathrm{T}}^{\mathrm{miss}}$.

7.5.3 Unclustered energy

The unclustered energy in each event comprises the low-p_{T} hadronic activity that is not clustered into a jet. Here, the scale of the unclustered energy is varied by $\pm 10\%$.

7.5.4 Pileup

The uncertainty in the number of pileup interactions in MC simulation stems from the luminosity in each bunch crossing and the effective inelastic cross section. In this analysis, the number of pileup interactions in MC is reweighted to match the data. The pileup uncertainty is evaluated by varying the effective inelastic cross section by $\pm 5\%$.

7.5.5 Hard scattering scale

The factorization scale, μ_F, determines the threshold separating the parton-parton hard scatter from softer interactions embodied in the parton distribution functions (PDFs). The renormalization scale, μ_R, sets the energy scale at which matrix element calculations occur. Both of these scales are set to be equal in the MC simulation. Here, they are varied simultaneously up and down by a factor of two to estimate the corresponding uncertainty.

7.5.6 Matching scale

The matrix element-parton shower matching threshold is used to interface the matrix elements generated in MadGraph with parton showers simulated in pythia. The value of this threshold is varied up and down by a factor of two.

7.5.7 Lepton energy scale

The electron energy scale is varied up and down by 0.6% in the ECAL Barrel ($\eta < 1.48$) and by 1.5% in the ECAL Endcap ($1.48 < \eta < 3$). The muon momentum scale is varied up and down by 0.2%. All variations are propagated into the $E_{\mathrm{T}}^{\mathrm{miss}}$.

7.5.8 Lepton identification and isolation

Event weights are applied to adjust the electron and muon yields in MC simulation to account for differences in the identification and isolation efficiencies between data and simulation. For muons, the uncertainty is taken to be 0.5% of the identification event weight, and 0.2% of the isolation event weight. For electrons, the uncertainties are estimated in bins of p_{T} and η, and are approximately $0.1 - 0.5\%$ of the combined event weight for identification and isolation [56].

7.5.9 b tagging efficiency

Event weights are applied to adjust the b jet yields in MC simulation to account for difference in the b-tagging efficiency between data and MC simulation [96]. The uncertainties are evaluated in bins of p_{T} and η.

7.5.10 Top quark p_{T} reweighting

Event weights are applied in order to compensate for a difference in the top quark p_{T} spectrum between data and MC simulation [97]. The uncertainty is evaluated by comparing the measurement in MC simulation with and without the weights applied. The event weights are not applied in the nominal result. This uncertainty is one-sided by construction.

7.5.11 Underlying event tunes and color reconnection

The underlying event tunes affect the modeling of soft hadronic activity that results from beam remnants and multiple-parton interactions in each event. The measurement is conducted with a $t\bar{t}$ sample from MC simulation using the 'Perugia 2011' tune, and compared to results using a sample with the 'Perugia 2011 mpiHi' and 'Perugia 2011 Tevatron' tunes [98] in PYTHIA. The largest difference is symmetrized to obtain the final uncertainty. The color reconnection (CR) uncertainty is evaluated by comparing measurement results using $t\bar{t}$ samples with the 'Perugia 2011' and 'Perugia 2011 no CR' tunes [98], where color reconnection effects are not included in the latter. The difference is symmetrized to obtain the final uncertainty.

7.5.12 Matrix element generator

The measurement is repeated using MC samples produced with the POWHEG event generator, which provides a next-to-leading order calculation of the $t\bar{t}$ production. These measurement results are compared with the reference $t\bar{t}$ MC sample, generated using MadGraph, to determine the corresponding uncertainty.

7.5.13 Parton distribution functions

The initial protons are described by parton distribution functions (PDFs). The corresponding uncertainty is evaluated by applying event weights in the MC simulation to reflect the CT10 PDF set [99] with 50 error eigenvectors. The total PDF uncertainty is determined by adding the variations corresponding to these error

sets in quadrature.

7.5.14 b quark fragmentation

The b quark fragmentation uncertainty includes two components that are implemented using event weights. The first component stems from the b quark fragmentation function, which can be tuned to the Lund fragmentation model in the PYTHIA Tune Z2* tune, or to empirical results from the ALEPH [100] and DELPHI [101] experiments. This component is evaluated by comparing the measurement results in MC simulation using these two tunes of the b quark fragmentation function, with the difference symmetrized to obtain the corresponding uncertainty. The second uncertainty component stems from the B hadron semi-leptonic branching fraction, which has an impact on the b quark JES due to the production of a neutrino. The corresponding uncertainty is evaluated by repeating the measurement with branching fraction values of 10.05% and 11.27%, which are variations from the nominal value of 10.50%. Both uncertainty components are combined in quadrature to obtain the total uncertainty.

7.6 Results and discussion

The results for each version of the likelihood fit are determined from 1 000 bootstrap pseudo-experiments in each fit, which are shown in Figs. 7.9 and 7.10. The 2D fit uses the $M_{b\ell}$ and M_{T2}^{bb} observables to simultaneously determine the values of M_t

Table 7.1: Systematic uncertainties for the 2D, 1D, hybrid, and MAOS likelihood fits. The ⊛ character highlights the uncertainty sources that are large in at least one of the likelihood fits.

	δM_t^{2D} (GeV)	δJSF^{2D}	δM_t^{1D} (GeV)	δM_t^{hyb} (GeV)	δM_t^{MAOS} (GeV)
JES (total) ⊛	$^{+0.06}_{-0.10}$	$^{+0.007}_{-0.006}$	$^{+0.54}_{-0.55}$	$^{+0.43}_{-0.46}$	$^{+0.65}_{-0.70}$
– MPF in situ	$^{+0.04}_{-0.04}$	$^{-0.002}_{+0.003}$	$^{-0.22}_{+0.21}$	$^{-0.18}_{+0.17}$	$^{-0.28}_{+0.24}$
– Intercalibration	$^{-0.01}_{+0.01}$	$^{-0.000}_{+0.000}$	$^{-0.04}_{+0.03}$	$^{-0.03}_{+0.03}$	$^{-0.04}_{+0.04}$
– Uncorrelated	$^{+0.04}_{-0.04}$	$^{-0.005}_{+0.005}$	$^{-0.39}_{+0.39}$	$^{-0.32}_{+0.31}$	$^{-0.47}_{+0.47}$
– Flavor	$^{+0.02}_{-0.09}$	$^{+0.004}_{-0.003}$	$^{+0.31}_{-0.32}$	$^{+0.25}_{-0.27}$	$^{+0.39}_{-0.43}$
b quark frag. (total) ⊛	$^{+0.39}_{-0.39}$	$^{+0.001}_{-0.001}$	$^{+0.40}_{-0.40}$	$^{+0.40}_{-0.40}$	$^{+0.67}_{-0.67}$
– Frag. function	$^{+0.38}_{-0.38}$	$^{+0.000}_{-0.000}$	$^{+0.38}_{-0.38}$	$^{+0.38}_{-0.38}$	$^{+0.64}_{-0.64}$
– Branching fraction	$^{+0.07}_{-0.07}$	$^{+0.001}_{-0.001}$	$^{+0.13}_{-0.13}$	$^{+0.12}_{-0.12}$	$^{+0.20}_{-0.20}$
JER	$^{-0.03}_{+0.08}$	$^{+0.001}_{-0.002}$	$^{+0.01}_{-0.05}$	$^{+0.00}_{-0.03}$	$^{+0.04}_{-0.04}$
Unclustered energy	$^{+0.10}_{-0.10}$	$^{+0.001}_{-0.001}$	$^{-0.02}_{+0.02}$	$^{-0.04}_{+0.01}$	$^{-0.11}_{+0.12}$
Pileup	$^{-0.06}_{+0.04}$	$^{-0.000}_{+0.000}$	$^{-0.06}_{+0.05}$	$^{-0.06}_{+0.05}$	$^{-0.06}_{+0.05}$
Electron energy scale	$^{-0.38}_{+0.39}$	$^{+0.002}_{-0.003}$	$^{-0.21}_{+0.21}$	$^{-0.24}_{+0.24}$	$^{-0.02}_{+0.05}$
Muon momentum scale	$^{-0.11}_{+0.09}$	$^{+0.001}_{-0.000}$	$^{-0.06}_{+0.05}$	$^{-0.07}_{+0.06}$	$^{-0.00}_{+0.01}$
Electron id/iso	$^{+0.07}_{-0.02}$	$^{-0.001}_{+0.000}$	$^{+0.03}_{-0.01}$	$^{+0.03}_{-0.01}$	$^{+0.01}_{-0.00}$
Muon id/iso	$^{+0.00}_{-0.00}$	$^{+0.000}_{-0.000}$	$^{+0.00}_{-0.00}$	$^{+0.00}_{-0.00}$	$^{+0.00}_{-0.00}$
b tagging	$^{+0.03}_{-0.03}$	$^{+0.000}_{-0.001}$	$^{-0.01}_{+0.01}$	$^{+0.00}_{-0.00}$	$^{+0.00}_{-0.00}$
Top quark p_T reweighting ⊛	$^{+0.93}_{-0.00}$	$^{-0.007}_{+0.000}$	$^{+0.40}_{-0.00}$	$^{+0.51}_{-0.00}$	$^{+0.72}_{-0.00}$
Hard scattering scale ⊛	$^{-0.36}_{+0.20}$	$^{+0.007}_{-0.003}$	$^{+0.31}_{-0.49}$	$^{+0.21}_{-0.47}$	$^{+0.33}_{-0.08}$
Matching scale ⊛	$^{-0.86}_{+0.30}$	$^{-0.004}_{+0.008}$	$^{-0.25}_{+0.11}$	$^{-0.37}_{+0.12}$	$^{+0.12}_{-0.12}$
Underlying event tunes ⊛	$^{+0.56}_{-0.56}$	$^{+0.007}_{-0.007}$	$^{+0.08}_{-0.08}$	$^{+0.11}_{-0.11}$	$^{+0.09}_{-0.09}$
Color reconnection	$^{+0.06}_{-0.06}$	$^{+0.001}_{-0.001}$	$^{+0.15}_{-0.15}$	$^{+0.13}_{-0.13}$	$^{+0.16}_{-0.16}$
ME Generator	$^{+0.18}_{-0.18}$	$^{-0.004}_{+0.002}$	$^{-0.19}_{+0.07}$	$^{-0.13}_{+0.07}$	$^{+0.11}_{-0.07}$
PDFs	$^{+0.14}_{-0.14}$	$^{+0.001}_{-0.001}$	$^{+0.17}_{-0.16}$	$^{+0.17}_{-0.15}$	$^{+0.17}_{-0.16}$
Total	$^{+1.31}_{-1.25}$	$^{+0.015}_{-0.014}$	$^{+0.91}_{-0.95}$	$^{+0.89}_{-0.93}$	$^{+1.27}_{-1.02}$

and JSF, yielding the results:

$$\mathrm{M_t^{2D}} = 171.56 \pm 0.46\,(\mathrm{stat})\,^{+1.31}_{-1.25}\,(\mathrm{syst})\,\mathrm{GeV},$$
$$\mathrm{JSF^{2D}} = 1.011 \pm 0.006\,(\mathrm{stat})\,^{+0.015}_{-0.014}\,(\mathrm{syst}).$$

The correlation between the $\mathrm{M_t}$ and JSF fit parameters in the 2D fit is shown in Fig. 7.11. The $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ distribution shapes corresponding to the fit results in a typical pseudo-experiment are shown in Fig. 7.12. The 2D fit is successful in mitigating the uncertainty due to the determination of JES, which is otherwise the largest source of systematic uncertainty in this measurement. In particular, this approach is robust against the flavor-dependent component of JES uncertainties – stemming from differences in the response between b jets, light quark jets, and gluon jets – since predominantly b jets are used for the determination of both $\mathrm{M_t}$ and JSF parameters. The underlying strategy, rooted in a simultaneous fit of two distributions with differing sensitivities to the JSF, does not rely on any specific assumptions about the event topology or final state. For this reason, it can be a viable option for JES uncertainty mitigation in a variety of physics scenarios.

The 1D fit is also based on the $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ observables, but constrains the JSF parameter to unity. The 1D fit gives:

$$\mathrm{M_t^{1D}} = 172.39 \pm 0.17\,(\mathrm{stat})\,^{+0.91}_{-0.95}\,(\mathrm{syst})\,\mathrm{GeV}.$$

In this approach, the JES accounts for the largest source of uncertainty. However, other uncertainties are reduced with respect to the 2D fit, resulting in an improved overall precision. The $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ distribution shapes corresponding to the 1D fit results in a typical pseudo-experiment are shown in Fig. 7.13.

The best overall precision is given by the hybrid fit, which is given by a linear combination of the 1D and 2D fit results. The 1D and 2D fits use the same set

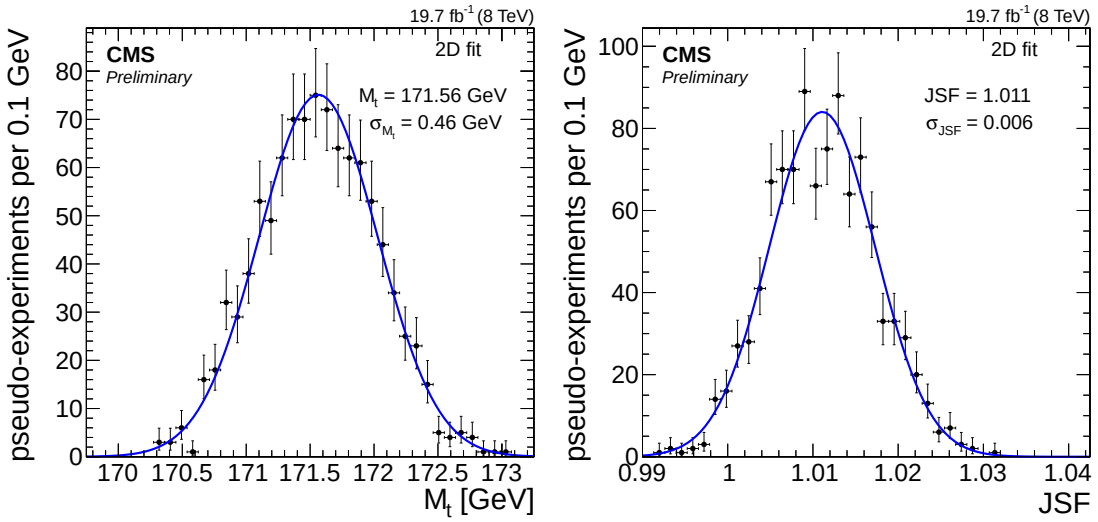

Figure 7.9: Likelihood fit results using 1k bootstrap pseudo-experiments for the (left) M_t and (right) JSF determined in the 2D fit.

of events and an identical likelihood function constructed from the $M_{b\ell}$ and M_{T2}^{bb} observables. These fits are fully correlated, with the only difference between them stemming from the treatment of the JSF parameter, which is fixed to unity in the 1D fit and acts as a free parameter in the 2D fit. The choice to fix the JSF parameter or allow it to float has an impact on the fit's sensitivity to a variety of uncertainty sources in addition to the JES. A linear combination of the 1D and 2D fits with $w_{hyb} = 0.8$ (Eq. (7.9)) achieves an optimal balance between all uncertainty sources, thus providing the best overall precision. The hybrid fit gives:

$$M_t^{hyb} = 172.22 \pm 0.18\,(stat)\,^{+0.89}_{-0.93}\,(syst)\,GeV.$$

The correlation between the M_t and JSF fit parameters in the hybrid fit is shown in Fig. 7.11.

The MAOS fit substitutes the $M_{b\ell}$ observable for an $M_{b\ell\nu}$ invariant mass, yielding a value of:

$$M_t^{MAOS} = 171.54 \pm 0.19\,(stat)\,^{+1.27}_{-1.02}\,(syst)\,GeV.$$

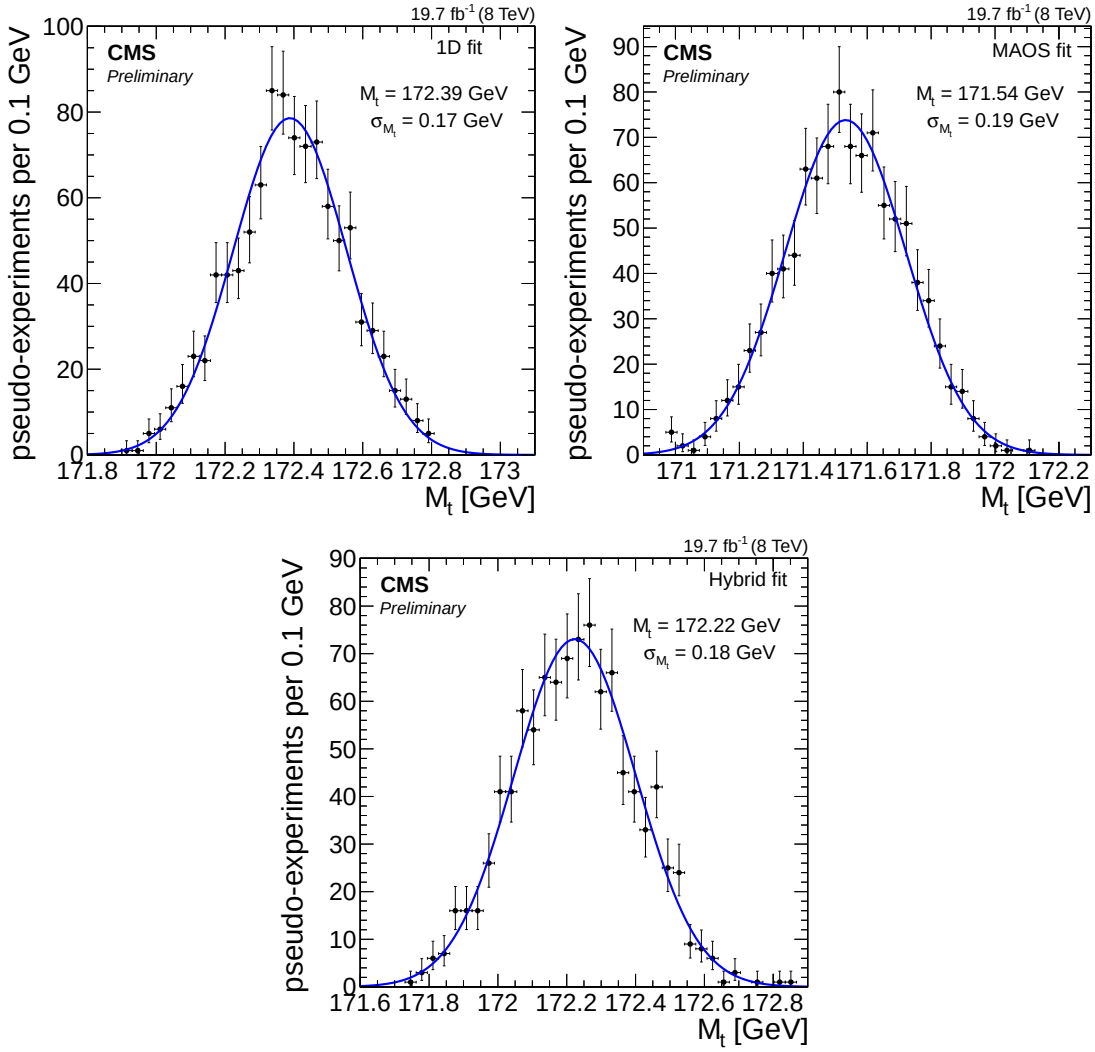

Figure 7.10: Likelihood fit results using 1k bootstrap pseudo-experiments for the (top left) 1D fit, (top right) MAOS fit, and (bottom) hybrid fit. The latter results are given by a linear combination of the 1D and 2D fits (Eq. (7.9)).

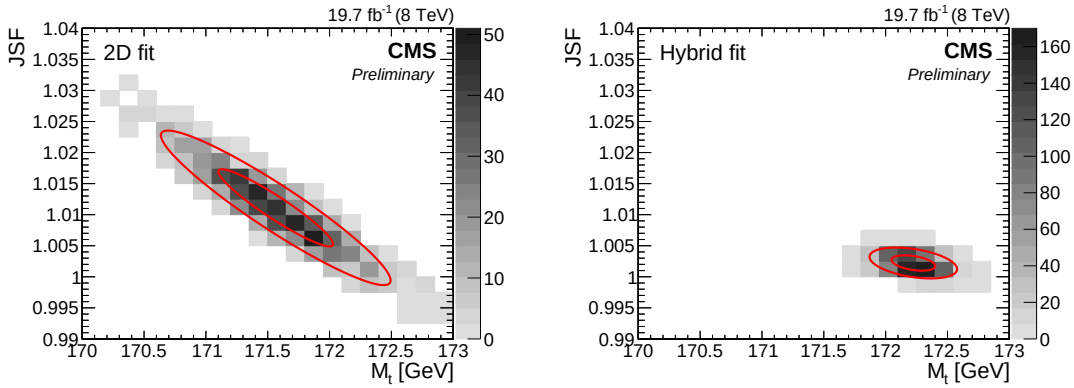

Figure 7.11: Likelihood fit results corresponding to the 2D fit (left) and hybrid fit (right), obtained using 1k pseudo-experiments constructed with the bootstrapping technique. The shaded gray histogram represents the number of pseudo-experiments in each bin of M_t and JSF. Two-dimensional contours corresponding to $-2\Delta\log(\mathcal{L}) = 1(4)$ are shown in red to indicate the one (two) sigma statistical intervals in M_t and JSF. The hybrid fit results are given by a linear combination of the 1D and 2D fit results using Eq. (7.9). The correlation coefficient between the M_t and JSF parameters is -0.94 in the 2D fit and -0.40 in the hybrid fit.

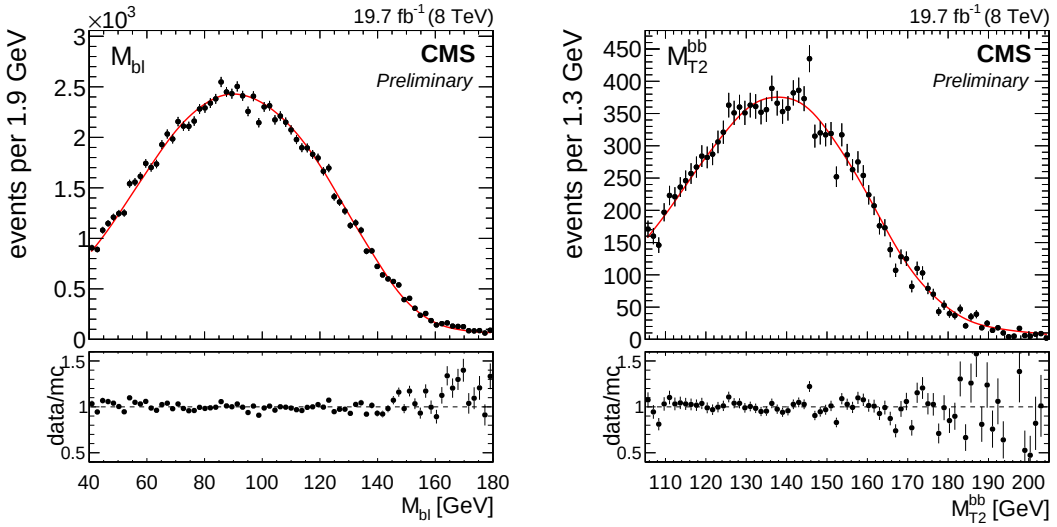

Figure 7.12: Maximum likelihood fit result in a typical pseudo-experiment of the 2D likelihood fit. The best-fit parameter values for this pseudo-experiment are $M_t = 171.99$ GeV and JSF $= 1.007$. When the JSF parameter is constrained to be unity in the 1D likelihood fit, the best-fit value of M_t is 172.48 GeV.

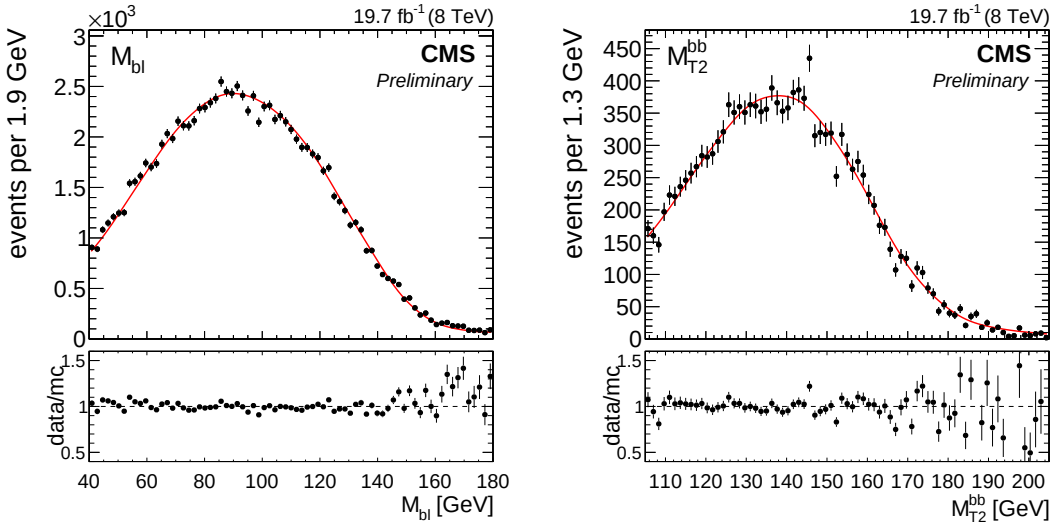

Figure 7.13: Maximum likelihood fit result in a typical pseudo-experiment of the 1D likelihood fit. The best-fit value of M_t for this pseudo-experiment is 172.48 GeV.

The MAOS observable presents a new approach for mass reconstruction in a decay topology characterized by underconstrained kinematics. Here, the MAOS fit provides a determination of M_t that is complementary to the 2D, 1D, and hybrid fits. The M_{T2}^{bb} and MAOS $M_{b\ell\nu}$ distribution shapes corresponding to the fit results in a typical pseudo-experiment are shown in Fig. 7.14.

The results for each version of the likelihood fit are summarized in Fig. 7.15.

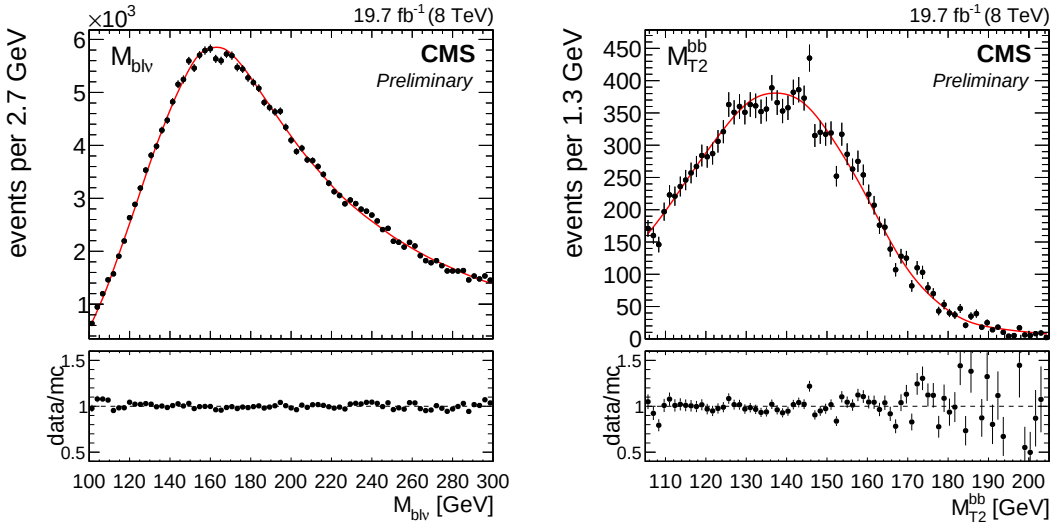

Figure 7.14: Maximum likelihood fit result in a typical pseudo-experiment of the MAOS likelihood fit. The best-fit value of M_t for this pseudo-experiment is 171.54 GeV.

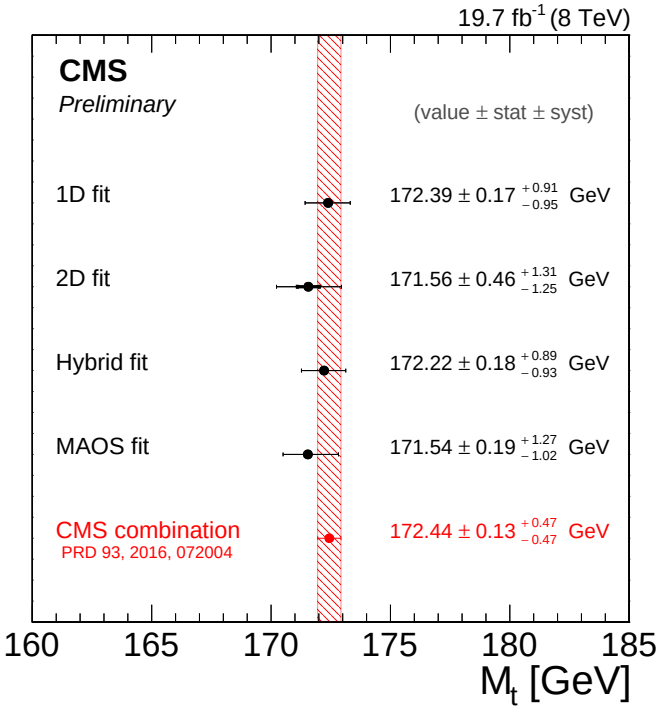

Figure 7.15: Summary of the 1D, 2D, hybrid, and MAOS likelihood fit results using the 2012 dataset corresponding to an integrated luminosity of $19.7 \pm 0.5\,\mathrm{fb}^{-1}$ and $\sqrt{s} = 8\,\mathrm{TeV}$. The most recent combination of $\mathrm{M_t}$ measurements by CMS [28] is shown for reference.

CHAPTER 8

SENSITIVITY OF KINEMATIC OBSERVABLES

Those who ignore Statistics are condemned to reinvent it.

– Bradley Efron

The measurement of the top quark mass presented in Chapter 7 is based on the $M_{b\ell}$, M_{T2}^{bb}, and MAOS $M_{b\ell\nu}$ observables. In the dileptonic $t\bar{t}$ decay topology, an event-by-event reconstruction of M_t is not possible due to a lack of kinematic constraints – these observables allow the value of M_t to be extracted from an ensemble of events, through features of their distribution shapes that are sensitive to M_t. These features include the kinematic endpoints of $M_{b\ell}$ and M_{T2}^{bb}, and the peak region of MAOS $M_{b\ell\nu}$, whose locations depend on the value of M_t (Section 7.2). In Figs. 7.2, 7.4, and 7.5, the distributions of these observables are shown to vary as a function of M_t^{MC}, thus providing sensitivity to the M_t parameter in a likelihood fit.

One way to quantify the sensitivity of these observables is given by the maximum likelihood (ML) method described in Appendix A. In the ML method, a likelihood function is constructed of the form:

$$\log \mathcal{L}(m) = \sum_{i}^{N} \log f(x_i|m), \tag{8.1}$$

where $f(x|m)$ is the distribution of observable x normalized to unity over its range, m is a free parameter, and N is the number of observations of x. (This likelihood has already been introduced in Eq. (7.7), where we have $x = M_{b\ell}$, M_{T2}^{bb}, or $M_{b\ell\nu}$, and $m = M_t$ or JSF.) Given this likelihood, the sensitivity of the observable x

is related to the curvature of $\mathcal{L}(m)$ in the vicinity of its maximum value – by Eq. (A.3), this curvature determines the statistical uncertainty, σ_m, corresponding to the best-fit value of m. In the ML approach, the value of σ_m obtained by the likelihood fit is a measure of sensitivity corresponding to the observable x.

In this chapter, we present a measure of sensitivity that is fundamental to the method described above. It is based on Fisher information [102, 103], described in Section 8.1 below, which determines the sensitivity of an observable x directly from its shape $f(x|m)$. It effectively estimates the curvature of $\mathcal{L}(m)$, as in Eq. (A.3), without requiring a full evaluation of the likelihood in Eq. (8.1). To illustrate why this is possible, we consider the sum in Eq. (8.1) over N observations of the observable x. Given a value of m, the distribution of x_i values entering the sum is already known – it is simply given by the shape $f(x|m)$. Thus, the sum in Eq. (8.1) can be converted to an integral of the form,

$$\sum_{i}^{N} \log f(x_i|m) \to N \int \mathrm{d}x\, f(x|m) \log f(x|m), \tag{8.2}$$

which is valid in the limit of large N. With the reliance on individual observations x_i now removed, the second derivatives in Eq. (A.3) can be evaluated directly from the shape using an integral over x. Section 8.2 features a case study, where the Fisher information method is demonstrated in the context of a previous top quark mass analysis from Ref. [29].

8.1 Fisher information

The Fisher information corresponding to the shape $f(x|m)$ is given by:

$$\mathcal{I}(m) = \int \left(\frac{\partial}{\partial m} \log f(x|m) \right)^2 f(x|m)\, \mathrm{d}x. \tag{8.3}$$

The quantity $\mathcal{I}(m)$ provides a measure of curvature near the maximum of the likelihood defined in Eq. (8.1). It can be interpreted as the variance of the slope, $(\partial \log f(x|m)/\partial m)$, known as the 'statistical score' of $f(x|m)$. The expression in Eq. (8.3) effectively combines the integral sketched in Eq. (8.2) with the second derivative of $\mathcal{L}(m)$ written in Eq. (A.3).

The Fisher information is related to the precision of a measurement by the Crámer-Rao (CR) bound:

$$\sigma_m^2 \geq \frac{1}{N \cdot \mathcal{I}(m)}, \tag{8.4}$$

where σ_m is the statistical uncertainty on parameter m. In a likelihood with large N, the shape of the likelihood near its maximum is roughly Gaussian, and the bound approaches an equality. This expression confirms the expected relationship $\sigma_m \propto 1/\sqrt{N}$ between the statistical uncertainty and the value of N, but also reveals the proportionality factor as the reciprocal of the Fisher information. It expresses the uncertainty σ_m in terms of the total number of events, the shape f, and the derivative $\partial f/\partial m$.

The Fisher information also provides a mathematical framework for quantifying the sensitivity of an observable at a specific point on its shape. In Chapter 7, the $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ observables have kinematic endpoints at approximately $\sqrt{\mathrm{M_t^2 - M_W^2}}$ and $\mathrm{M_t}$, respectively; the MAOS $\mathrm{M_{b\ell\nu}}$ observable is an invariant mass whose shape contains a peak near the value of $\mathrm{M_t}$. Because these features carry a dependence on the value of $\mathrm{M_t}$, the endpoint regions of $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$ and the peak region of $\mathrm{M_{b\ell\nu}}$ are expected to contribute significantly to the sensitivity of these observables. To relate these local features to the Fisher information, we consider the integral in Eq. (8.3) over the value of observable x. Here, the *integrand* of the Fisher information can be interpreted as the contribution to the total sensitivity stemming

from a specific value of x. Rewriting the integrand in a more convenient form, we define the 'local shape sensitivity' function by:

$$s(x|m) \equiv \frac{1}{f(x|m)} \left[\frac{\partial f(x|m)}{\partial m} \right]^2 \tag{8.5}$$

This function is also known as the Fisher information density. It is shown for the $\mathrm{M_{b\ell}}$, $\mathrm{M_{T2}^{bb}}$, and MAOS $\mathrm{M_{b\ell\nu}}$ observables by the red curves in Figs. 7.2, 7.4, and 7.5, with $m = \mathrm{M_t}$ and the JSF parameter fixed to unity. It is observed to peak near the kinematic endpoints of $\mathrm{M_{b\ell}}$ and $\mathrm{M_{T2}^{bb}}$, and on the left-side edge of $\mathrm{M_{b\ell\nu}}$. The values of x where $s(x|m) = 0$ coincide with the stationary points at which the gray shapes in Figs. 7.2, 7.4, and 7.5 intersect. This is a reflection of the fact that in a likelihood fit, events with a value of x near a stationary point make little or no contribution to the determination of m. In general, the shape of $s(x|m)$ for each observable establishes a link between the underlying kinematic properties of the observable and regions of high and low sensitivity on its shape. In this analysis, it provides heuristic information about the $\mathrm{M_{b\ell}}$, $\mathrm{M_{T2}^{bb}}$, and MAOS $\mathrm{M_{b\ell\nu}}$ distributions, and their sensitivity to the value of $\mathrm{M_t}$.

In addition to providing heuristic information, the local shape sensitivity function was used in the top quark mass measurement to identify potential overfitting effects in the GP shapes. Overfitting occurs when the interpolation between GP training points is not smooth, causing fluctuations in the shape that may be difficult to identify by eye. Such fluctuations can be a source of bias, both in the determination of $\mathrm{M_t}$ and its corresponding uncertainties. A typical symptom of overfitting is an under-estimated statistical uncertainty on the value of $\mathrm{M_t}$. This can occur when fluctuations in the GP shape increase the value of the slope $\partial f(x|\mathrm{M_t})/\partial \mathrm{M_t}$, thus artificially increasing the Fisher information of the corresponding shape (Eq. (8.3)). The issue is easily revealed by the shape of $s(x|\mathrm{M_t})$, which acquires visible fluctuations when overfitting is indeed present. In such cases,

overfitting can be mitigated by increasing relevant GP hyperparameter values to improve the smoothness of the GP shape.

8.2 Case study: sensitivity of the endpoint method

In a mass measurement based on the 7 TeV dataset, the value of M_t is determined using an endpoint fit to the $M_{b\ell}$ and M_{T2} distributions [29]. In this method, the endpoint regions of the distributions are assumed to be well-approximated by a first-order polynomial descending to zero at the location of the endpoint (i.e. a 'kinked line'). A likelihood fit restricted to the endpoint regions of $M_{b\ell}$ and M_{T2} is performed, with reconstructed object resolutions convolved with the kinked-line approximation in each event. The background due to light quark and gluon jets mistagged as b jets is modeled using a control region extracted from data. The results of the fit are shown in Fig. 8.1. The value of M_t determined using this method is minimally-dependent on MC simulation – the method does not rely on simulation to estimate distribution shapes, nor does it perform a direct calibration to the M_t^{MC} parameter in simulation.

Although the endpoint regions of $M_{b\ell}$ and M_{T2} contain most of the sensitivity in these distributions, a fraction of the total available statistical precision is lost due to the restrictive fit range of the endpoint fit. The precision that is lost is not evaluated in Ref. [29], since that would require a comparison to a likelihood fit based on the distribution shapes over their full range. The likelihood used in Ref. [29] can only accommodate a kinked-line shape, however, and cannot be extended to include the full shapes.

In this section, we compare the precision of the endpoint method with that of

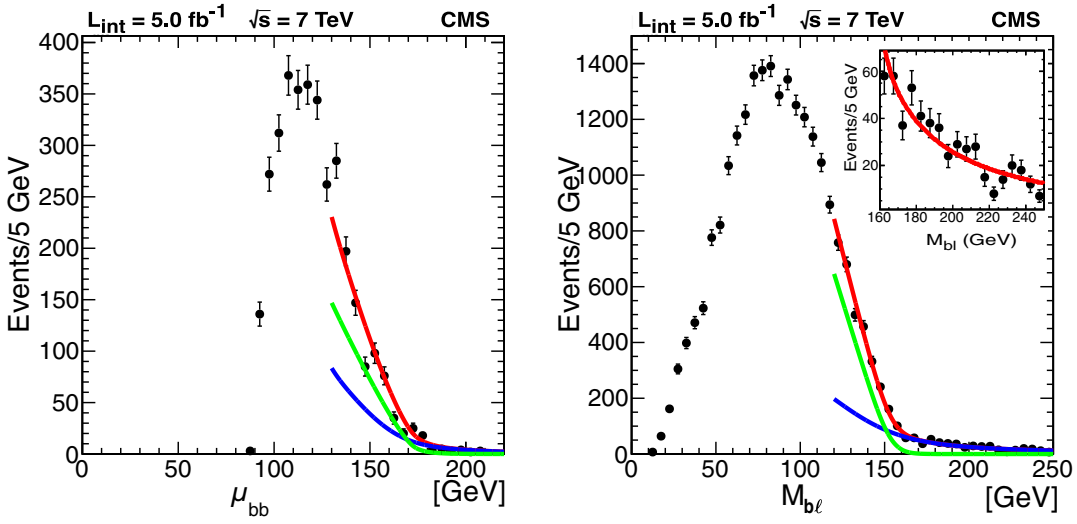

Figure 8.1: Results of the fit for M_t in Ref. [29]. The upper red line corresponds to the full fit, while the green (middle) and blue (lowest) curves are for the signal and background shapes, respectively.

a likelihood fit to the full distribution shapes. Our method, based on the local sensitivity function, does not require the evaluation of a likelihood, instead using only knowledge of the distribution shapes and their evolution in the M_t parameter. Given this information, an integral of the local sensitivity function over the relevant fit range provides an estimate of the corresponding precision.

To find the dependence of the statistical precision, given by $1/\sigma^2_{\rm tot}$, on the fit range of a kinematic distribution, we first need to address the distribution's normalization. Taking $[a, b]$ to be the full range of a distribution, a limited range given by $[y, b]$ scales the CR bound, and thus the precision, by a factor:

$$\mathcal{S}(y) = \frac{\sigma^2_{\rm tot}}{\sigma^2_y} = \frac{N_y}{N_{\rm tot}} \frac{\mathcal{I}'(\theta; y, b)}{\mathcal{I}(\theta; a, b)}, \tag{8.6}$$

where σ_y is the statistical uncertainty of a measurement with fit range $[y, b]$, and $\mathcal{I}'$ is the Fisher information corresponding to the reduced fit range. The latter is

given by:

$$\mathcal{I}'(\theta; y, b) = \int_y^b \left(\frac{\partial}{\partial \theta} \log f'(x|\theta) \right)^2 f'(x|\theta)\, \mathrm{d}x, \tag{8.7}$$

where $f'(x|\theta)$ is the distribution shape normalized to unity on the fit range $[y, b]$:

$$\int_y^b f'(x|\theta)\, \mathrm{d}x = 1 = \frac{1}{k} \int_y^b f(x|\theta)\, \mathrm{d}x. \tag{8.8}$$

Here the original shape, $f(x|\theta)$, is normalized to unity on the full range, $[a, b]$, so that its normalization on $[y, b]$ is given by the factor $k \leq 1$. Writing $f'(x|\theta) = (1/k) f(x|\theta)$, we have:

$$\mathcal{I}'(\theta; y, b) = \frac{1}{k} \mathcal{I}(\theta; y, b). \tag{8.9}$$

The factor k can also be expressed in terms of the total number of events – in the limit of large N_y, the normalization condition in Eq. (8.8) gives:

$$k = \frac{N_y}{N_{\mathrm{tot}}}, \tag{8.10}$$

so that fraction of total precision is given by:

$$\mathcal{S}(y) = \frac{\mathcal{I}(\theta; y, b)}{\mathcal{I}(\theta; a, b)} = \frac{1}{\mathcal{I}(\theta; a, b)} \int_y^b \left(\frac{\partial}{\partial \theta} \log f(x|\theta) \right)^2 f(x|\theta)\, \mathrm{d}x. \tag{8.11}$$

The function $\mathcal{S}(y)$ is also known as the cumulative distribution function (CDF) of $s(x|m)$.

The function $\mathcal{S}(y)$ for the $\mathrm{M}_{\mathrm{b}\ell}$ distribution is represented by the blue line in Fig. 8.2 . It is evident that most of the sensitivity is in the endpoint region of the distribution. For a fit range approximately corresponding to the one used in Ref. [29], with $y = 120\,\mathrm{GeV}$, one obtains a value of $\mathcal{S} \sim 0.3$. This indicates that the endpoint method, despite using only $\sim 20\%$ of the $\mathrm{M}_{\mathrm{b}\ell}$ values available in the dataset, retains roughly 70% of the total available statistical precision.

The conclusions of this study are approximate, since the fit in Ref. [29] has additional free parameters controlling the normalization of data-driven background

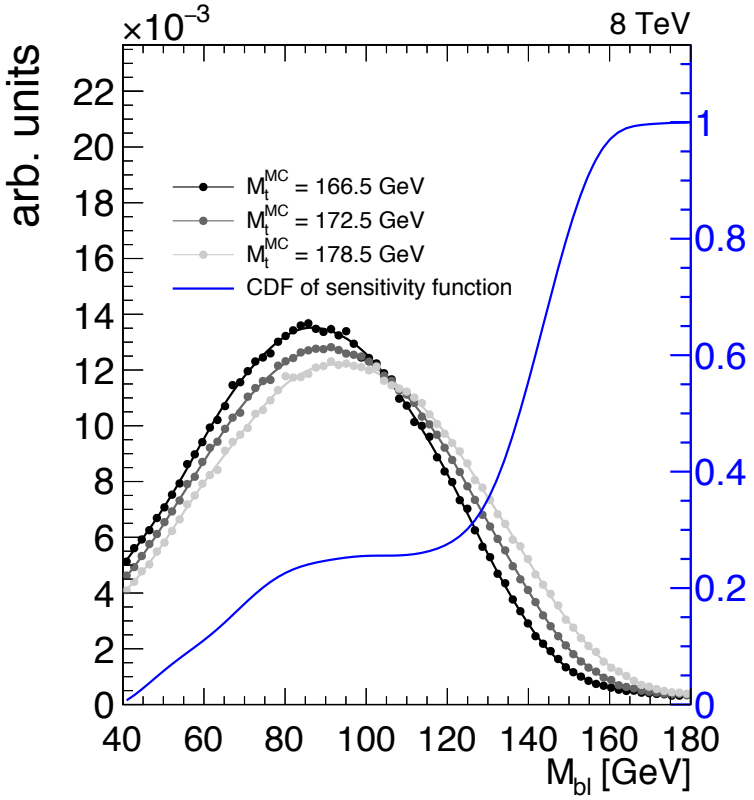

Figure 8.2: The $M_{b\ell}$ distribution for three values of M_t^{MC} is shown, along with a curve corresponding to $\mathcal{S}(y)$, the CDF of the local shape sensitivity function, in blue.

shapes. For a more complete treatment, the Fisher information can be computed for multiple variables, yielding a Fisher information matrix. This matrix accounts for the sensitivity corresponding to each variable, as well as correlations between the variables. Such a multivariable approach can also be used to gain insight on the two-parameter fit for the M_t and JSF parameters in Chapter 7. The Fisher information matrix remains to be explored in future studies.

CHAPTER 9

SUMMARY AND CONCLUSIONS

Equipped with his five senses, man explores the universe around him and calls the adventure Science.

– Edwin Hubble

More than twenty years after the discovery of the top quark, measurements at the LHC continue to refine our understanding of its properties. This dissertation presents a measurement of the top quark mass in the dileptonic $\mathrm{t\bar{t}}$ decay channel using proton-proton collision data at $\sqrt{s} = 8\,\mathrm{TeV}$ recorded by the CMS experiment at the LHC. The dataset was collected in 2012 and corresponds to an integrated luminosity of $19.7 \pm 0.5\,\mathrm{fb}^{-1}$. The top quark mass analysis is based on the $\mathrm{M_{b\ell}}$, $\mathrm{M_{T2}^{bb}}$, and MAOS $\mathrm{M_{b\ell\nu}}$ observables, which allow for mass reconstruction in decay topologies that are kinematically underconstrained. These observables are employed in several versions of an event-by-event likelihood fit, with the results summarized in Fig. 7.15. The 2D fit provides a measurement of $\mathrm{M_t}$ that is robust against uncertainties due to JES determination, yielding $\mathrm{M_t^{2D}} = 171.56 \pm 0.46\,(\mathrm{stat})\,^{+1.31}_{-1.25}\,(\mathrm{syst})\,\mathrm{GeV}$ and $\mathrm{JSF^{2D}} = 1.011 \pm 0.006\,(\mathrm{stat})\,^{+0.015}_{-0.014}\,(\mathrm{syst})$. The most precise measurement of the top quark mass is achieved by the hybrid fit, which gives $\mathrm{M_t^{hyb}} = 172.22 \pm 0.18\,(\mathrm{stat})\,^{+0.89}_{-0.93}\,(\mathrm{syst})\,\mathrm{GeV}$.

The JES calibration technique employed in the 2D fit takes a novel approach that can be implemented in the dileptonic $\mathrm{t\bar{t}}$ decay topology. Standard methods for JES calibration in top quark analyses require the reconstruction of a $\mathrm{W} \to \mathrm{jj}$

decay, which is available only in the all-hadronic and semi-leptonic channels. In such methods, the dijet invariant mass is calibrated to the known W boson mass, with the corresponding calibration factor also applied to the b jets in the event. The method featured here does not rely on a hadronic W boson decay, instead achieving sensitivity to the JES through a simultaneous fit to the $M_{b\ell}$ and M_{T2}^{bb} distributions. Because this approach targets the b jets directly, without relying on light quark jets to extract the relevant calibration factor, it is robust against flavor-dependent JES uncertainties stemming from differences between the b jet and light quark jet response.

A GP regression technique used in the top quark mass analysis models the shapes of the $M_{b\ell}$, M_{T2}^{bb}, and MAOS $M_{b\ell\nu}$ distributions, as well as their evolution in the M_t and JSF parameters. The resulting shapes are non-parametric, and therefore largely model-independent. When employed for likelihood fitting, the GP shapes provide unbiased results over a range of M_t and JSF values. The GP shapes in this analysis are trained using bins of kinematic distributions, along with a Gaussian covariance function that specifies the smoothness properties of each shape. Overall, the GP regression technique has proven to be a powerful alternative to the currently established set of interpolation methods used in high energy physics analysis.

The sensitivity of the kinematic observables to the value of M_t is investigated using a local shape sensitivity function, also known as the Fisher information density. This function conveys the sensitivity of an observable at each point on its shape. It highlights the endpoint regions of the $M_{b\ell}$ and M_{T2}^{bb} distributions, which contain most of the sensitivity in these observables, and the left-side edge of MAOS $M_{b\ell\nu}$. The integral of the local shape sensitivity function provides an estimate of

the total statistical uncertainty in a measurement of M_t. In the top quark mass analysis, the local shape sensitivity function is used for fit range optimization, and as a cross-check for overfitting in the GP shapes. It also provides important heuristic information on the kinematic observables used in the analysis.

This dissertation also presents a $\vec{\not{E}}_T$ significance variable, which estimates the per-event $\vec{\not{E}}_T$ resolution by combining individual JERs with a parametrized estimate of the unclustered energy resolution. The significance variable is tuned using data-driven techniques, and has been studied as a cut variable for background rejection in the $W \to e\nu$ channel. Because event topology can have a large impact on the magnitude and direction of $\vec{\not{E}}_T$ smearing in an event, the $\vec{\not{E}}_T$ significance variable has been shown to outperform simpler alternatives, such as $\not{E}_T$ or $\not{E}_T/\sqrt{\sum E_T}$, in discriminating $W \to e\nu$ events from a QCD multijet background. The $\vec{\not{E}}_T$ significance variable has been commissioned for Run 2 analyses, and can be introduced as a cut variable to extend the reach of BSM physics searches in scenarios containing high-p_T jet activity or low $\not{E}_T$.

In the last several years, the LHC and Tevatron experiments have achieved a sub-GeV precision on the top quark mass. Recent LHC and world combinations of top quark mass measurements are shown in Fig. 9.1. The most precise combination includes measurements in the dileptonic, semi-leptonic, and all-hadronic channels at CMS, achieving a precision of 0.48 GeV [28]. These measurements are limited by their systematic uncertainties, stemming mostly from JES determination and the modeling of $t\bar{t}$ events in simulation [30]. Progress in understanding these sources of uncertainty can lead to further improvements in the determination of M_t. A projection of the precision achievable in future top quark mass measurements at CMS is shown in Fig. 9.2. The projection includes standard mass measurement

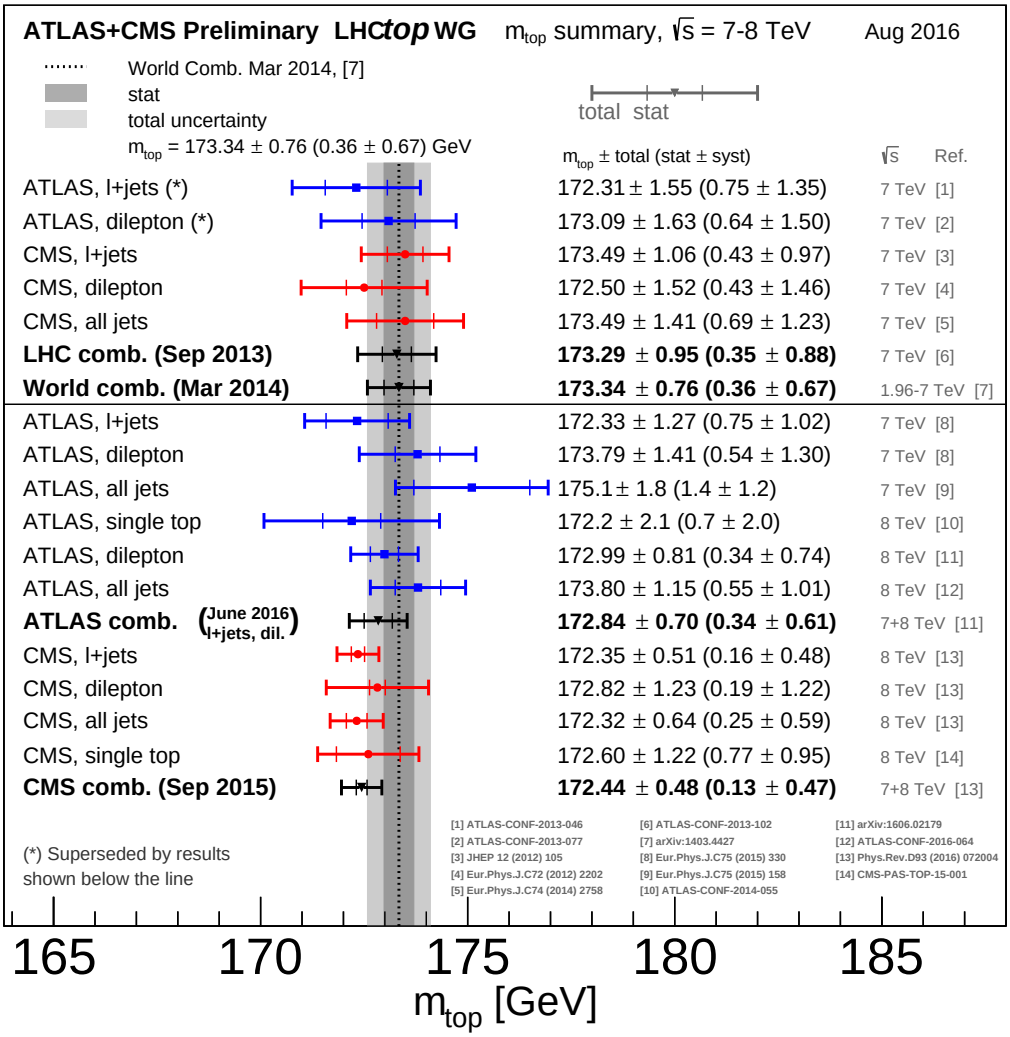

Figure 9.1: World and LHC top quark mass combinations [4].

techniques, as well as alternatives methods based on kinematic endpoints, the B hadron decay length (L_{xy}), and the reconstruction of J/ψ mesons stemming from B hadron decays. In conventional methods, where a full kinematic reconstruction of the $t\bar{t}$ system is conducted, uncertainties due to the determination of the JES can be reduced by improving jet and detector calibrations, and employing *in situ* JES calibration techniques that are robust against flavor-dependent effects. One such technique is introduced in this dissertation. Uncertainties stemming from

parameters in MC simulation – such as the renormalization and factorization scales, the matrix element-parton shower matching threshold, and the modeling of color reconnection – can be reduced through further study of $t\bar{t}$ events at the particle level, as well as differential studies of cross sections and top quark mass observables. Theoretical progress is also needed to understand the connection between M_t^{MC} and a well-defined top quark mass, such as the top quark pole mass or $\overline{MS}$ mass.

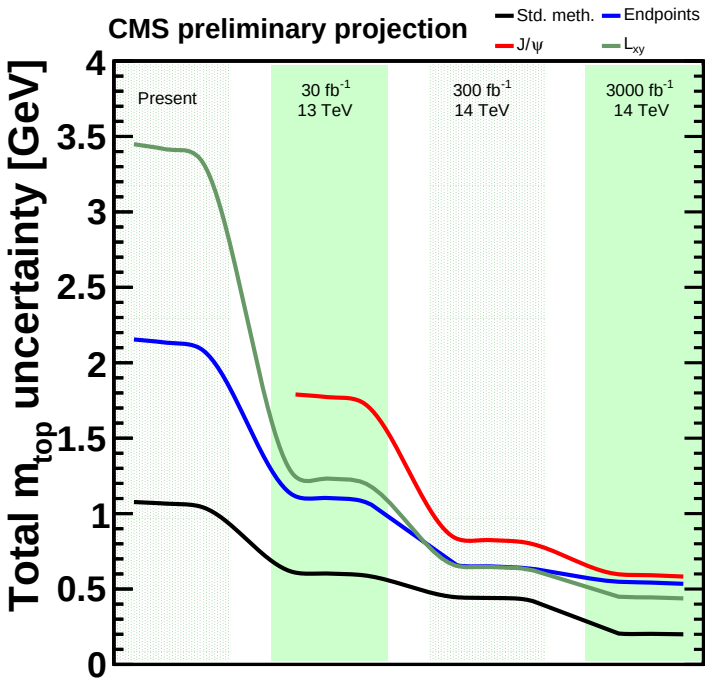

Figure 9.2: Projection of precision achievable in future top quark mass analyses at CMS [30].

Although the precision achievable on M_t at a hadron collider is ultimately limited by nonperturbative effects in QCD, significant improvements to the precision can be attained at a future high energy lepton collider [104, 105]. At a lepton collider, a threshold scan can be conducted to precisely determine the position of the peak in the $t\bar{t}$ production cross section, with the location of the peak directly related to the value of M_t. Current estimates give a precision better than 50 MeV

on the value of M_t in such measurements. In addition to measurements of the top quark, a future lepton collider can provide high-precision studies of the W, Z, and Higgs bosons, with uncertainties at the level of one part per mill [106]. Such a program would test the self-consistency of the SM in a regime not accessible by the LHC, and set additional constraints on possible models of BSM physics.

The coming decades may shed light on a number of open questions in particle physics: the nature of dark matter; the origin of baryon asymmetry in the observable universe; the mechanism responsible for nonzero neutrino masses; the nature of CP-violation and its apparent absence in QCD; and a resolution to the SM hierarchy problem. These topics are at the forefront of fundamental physics, with implications spanning orders of magnitude, from the interactions of subatomic particles at 10^{-18} m to the birth and evolution of the observable universe. The potential for exploration and discovery is great.

APPENDIX A

THE MAXIMUM LIKELIHOOD METHOD

In the maximum likelihood (ML) method, we construct a likelihood of the form:

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i}^{N} f(x_i|\boldsymbol{\theta}), \tag{A.1}$$

where $f(x|\boldsymbol{\theta})$ is the distribution, or probability density function, corresponding to an observable x and set of parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$, and N is the number of observations of x [34]. The distribution $f(x|\boldsymbol{\theta})$ is normalized to unity over its range. (In the measurement outlined in Chapter 7, we have $x = \mathrm{M}_{\mathrm{b}\ell}$, $\mathrm{M}_{\mathrm{T2}}^{\mathrm{bb}}$, or $\mathrm{M}_{\mathrm{b}\ell\nu}$, and $\boldsymbol{\theta} = (\mathrm{M_t}, \mathrm{JSF})$.) This likelihood can be interpreted as the probability of observing the values $\mathbf{x} = (x_1, \ldots, x_N)$, given a set of parameter values $\boldsymbol{\theta}$, or $\mathcal{L}(\boldsymbol{\theta}) = P(\mathbf{x}|\boldsymbol{\theta})$. Here, there is an implicit assumption that the values of x are independent and identically distributed (IID), so that the total probability can be written as a factorized product over N individual observations. Minor correlations may exist between values of x, especially if more than one value is extracted from a single event, but they can typically be treated as negligible.

Given the likelihood in Eq. (A.1), the parameter values $\boldsymbol{\theta}$ are estimated by maximizing $\mathcal{L}(\boldsymbol{\theta})$. The resulting ML estimators ($\hat{\boldsymbol{\theta}}$) solve the likelihood equations:

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = 0, \tag{A.2}$$

for $i = 1, \ldots, k$. In practice, the analytical form of $\mathcal{L}(\boldsymbol{\theta})$ is usually not known, and the solution must be found numerically. The analyses described in Chapters 5 and 7 use the MINUIT numerical minimization program as implemented in the ROOT software package [107, 108].

The standard deviation, or statistical uncertainty, of the values $\hat{\boldsymbol{\theta}}$ is determined from the shape of the likelihood $\mathcal{L}(\boldsymbol{\theta})$ in the vicinity of its maximum. Under certain regularity conditions, the likelihood has the property of *asymptotic normality*, ensuring that its shape is approximately Gaussian for small $|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}|$ or large N. By the Lindeberg-Lévy central limit theorem, the corresponding covariance matrix, $V_{ij} = \mathrm{cov}[\hat{\theta}_i, \hat{\theta}_j]$, can be written:

$$(\hat{V}^{-1})_{ij} = \frac{\partial^2 \log \mathcal{L}}{\partial\theta_i \partial\theta_j}, \Bigg|_{\hat{\boldsymbol{\theta}}} . \tag{A.3}$$

Numerically, the s-standard-deviation errors on the values $\hat{\boldsymbol{\theta}}$ can be determined by the hypersurface given by the $\boldsymbol{\theta}$ satisfying:

$$\log \mathcal{L}(\boldsymbol{\theta}) = \log \mathcal{L}(\hat{\boldsymbol{\theta}}) - s^2/2. \tag{A.4}$$

For each θ_i, its minimum and maximum values that lie on this hypersurface correspond to the s-standard-deviation error on the parameter.

BIBLIOGRAPHY

[1] O. S. Brning, P. Collier, P. Lebrun, S. Myers, R. Ostojic, J. Poole, and P. Proudlock, *LHC Design Report*. CERN, Geneva, 2004. https://cds.cern.ch/record/782076.

[2] W. de Boer, "The Discovery of the Higgs Boson with the CMS Detector and its Implications for Supersymmetry and Cosmology," in *Time and Matter 2013 (TAM2013) Venice, Italy*. 2013. arXiv:1309.0721 [hep-ph].

[3] CMS Collaboration, "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV," *Eur. Phys. J. C* **75** no. 5, (2015) 212, arXiv:1412.8662 [hep-ex].

[4] LHC Top Physics Working Group, "LHCtopWG Summary Plots." https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots.

[5] D0 Collaboration, "Useful diagrams of top signals and backgrounds." https://www-d0.fnal.gov/Run2Physics/top.

[6] M. Baak, J. Cth, J. Haller, A. Hoecker, R. Kogler, K. Mnig, M. Schott, and J. Stelzer, "The global electroweak fit at NNLO and prospects for the LHC and ILC," *Eur. Phys. J. C* **74** (2014) 3046, arXiv:1407.3792 [hep-ph].

[7] G. Degrassi, S. Di Vita, J. Elias-Miro, J. R. Espinosa, G. F. Giudice, G. Isidori, and A. Strumia, "Higgs mass and vacuum stability in the Standard Model at NNLO," *JHEP* **08** (2012) 098, arXiv:1205.6497 [hep-ph].

[8] M. C. Smith and S. S. Willenbrock, "Top quark pole mass," *Phys. Rev. Lett.* **79** (1997) 3825–3828, arXiv:hep-ph/9612329 [hep-ph].

[9] A. Harmon, "Sherpa and open science grid: Predicting the emergence of jets." https://sciencenode.org/feature/sherpa-and-open-science-grid-predicting-emergence-jets.php.

[10] C. Lefevre, "LHC: the guide (English version). Guide du LHC (version anglaise)." Feb, 2009.

[11] "LHC machine outreach." https://lhc-machine-outreach.web.cern.ch/lhc-machine-outreach/lhc_in_pictures.htm.

[12] "CMS detector drawings." https://cms-docdb.cern.ch/cgi-bin/PublicDocDB/ShowDocument?docid=5955.

[13] "CMS photos." https://cds.cern.ch/collection/CMS%20Photos?ln=en.

[14] CMS Collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker," *JINST* **9** no. 10, (2014) P10009, arXiv:1405.6569 [physics.ins-det].

[15] W. Erdmann, "The CMS pixel detector," *International Journal of Modern Physics A* **25** no. 07, (2010) 1315–1337, http://www.worldscientific.com/doi/pdf/10.1142/S0217751X10049098.

[16] CMS Collaboration, "CMS, tracker technical design report,". http://weblib.cern.ch/abstract?CERN-LHCC-98-06.

[17] A. Ryd, S. Stroiney, S. Das, K. Ecklund, J. Thompson, and B. Kreis, "CMS pixel online software and calibrations." 2013.

[18] A. Dominguez *et al.*, "CMS Technical Design Report for the Pixel Detector Upgrade,". https://cds.cern.ch/record/1481838.

[19] CMS Collaboration, "Performance and Operation of the CMS Electromagnetic Calorimeter," *JINST* **5** (2010) T03010, arXiv:0910.3423 [physics.ins-det].

[20] "CMS detector layout." http://www.hephy.at/user/friedl/diss/html/node8.html.

[21] CMS Collaboration, "The CMS experiment at the CERN LHC," *JINST* **3** (2008) S08004.

[22] CMS Collaboration, *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical Design Report CMS. http://cds.cern.ch/record/706847.

[23] S. Cittolin, A. Rcz, and P. Sphicas, *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. CERN, Geneva, 2002. http://cds.cern.ch/record/578006.

[24] CMS Collaboration, "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV," *Submitted to: JINST* (2016) , arXiv:1607.03663 [hep-ex].

[25] CMS Collaboration, "Identification of b-quark jets with the CMS experiment," *JINST* **8** (2013) P04013, arXiv:1211.4462 [hep-ex].

[26] CMS Collaboration, "Performance of the CMS missing transverse momentum reconstruction in pp data at $\sqrt{s} = 8$ TeV," *JINST* **10** no. 02, (2015) P02006, arXiv:1411.0511 [physics.ins-det].

[27] N. Mirman, Y. Wang, and J. Alexander (for the CMS Collaboration), "Missing transverse energy significance at CMS," in *Proceedings, 2nd Conference on Large Hadron Collider Physics Conference (LHCP 2014): New York, USA, June 2-7, 2014*. 2014. arXiv:1409.3028 [hep-ex]. https://inspirehep.net/record/1315814/files/arXiv:1409.3028.pdf.

[28] CMS Collaboration, "Measurement of the top quark mass using proton-proton data at $\sqrt{s} = 7$ and 8 TeV," *Phys. Rev. D* **93** (2016) 072004, arXiv:1509.04044 [hep-ex].

[29] S. Chatrchyan *et al.*, "Measurement of masses in the $t\bar{t}$ system by kinematic endpoints in pp collisions at $\sqrt{s} = 7$ TeV," *Eur. Phys. J. C* **73** (2013) 2494, arXiv:1304.5783 [hep-ex].

[30] CMS Collaboration, "Projected improvement of the accuracy of top-quark mass measurements at the upgraded LHC,".

[31] H. Baer and X. Tata, *Weak Scale Supersymmetry*. Cambridge University Press, 2006.

[32] P. Langacker, *The Standard Model and Beyond*. CRC Press, 2010.

[33] M. Peskin and D. Shroeder, *An Introduction to Quantum Field Theory*. Westview Press, 1995.

[34] K. A. Olive *et al.*, "Review of Particle Physics," *Chin. Phys. C* **38** (2014) 090001. http://pdg.lbl.gov.

[35] M. Kobayashi and T. Maskawa, "CP-Violation in the Renormalizable Theory of Weak Interaction," *Progress of Theoretical Physics* **49** no. 2, (1973) 652–657, http://ptp.oxfordjournals.org/content/49/2/652.full.pdf+html. http://ptp.oxfordjournals.org/content/49/2/652.abstract.

[36] CDF Collaboration, "Evidence for top quark production in $p\overline{p}$ collisions at $\sqrt{}$ s =1.8 TeV," *Phys. Rev. Lett.* **73** (Jul, 1994) 225–231. http://link.aps.org/doi/10.1103/PhysRevLett.73.225.

[37] CDF Collaboration, "Observation of top quark production in $\overline{p}p$ collisions with the collider detector at fermilab," *Phys. Rev. Lett.* **74** (Apr, 1995) 2626–2631. http://link.aps.org/doi/10.1103/PhysRevLett.74.2626.

[38] D0 Collaboration, "Observation of the top quark," *Phys. Rev. Lett.* **74** (Apr, 1995) 2632–2637. http://link.aps.org/doi/10.1103/PhysRevLett.74.2632.

[39] M. Czakon, P. Fiedler, and A. Mitov, "Total top-quark pair-production cross section at hadron colliders through $\mathcal{O}(\alpha_S^4)$," *Phys. Rev. Lett.* **110** (Jun, 2013) 252004. http://link.aps.org/doi/10.1103/PhysRevLett.110.252004.

[40] ATLAS Collaboration, "Differential top-antitop cross-section measurements as a function of observables constructed from final-state particles using pp collisions at $\sqrt{s} = 7$ TeV in the ATLAS detector," *JHEP* **06** (2015) 100, arXiv:1502.05923 [hep-ex].

[41] CMS Collaboration, "Measurement of the $\mathrm{t\bar{t}}$ production cross section in the all-jets final state in pp collisions at $\sqrt{s} = 8$ TeV," *Eur. Phys. J. C* **76** no. 3, (2016) 128, arXiv:1509.06076 [hep-ex].

[42] N. Kidonakis, "Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production," *Phys. Rev. D* **83** (May, 2011) 091503. http://link.aps.org/doi/10.1103/PhysRevD.83.091503.

[43] N. Kidonakis, "NNLL resummation for s-channel single top quark production," *Phys. Rev. D* **81** (2010) 054028, arXiv:1001.5034 [hep-ph].

[44] N. Kidonakis, "Two-loop soft anomalous dimensions for single top quark associated production with a W- or H-," *Phys. Rev. D* **82** (2010) 054018, arXiv:1005.4451 [hep-ph].

[45] C. Campagnari and M. Franklin, "The Discovery of the top quark," *Rev. Mod. Phys.* **69** (1997) 137–212, arXiv:hep-ex/9608003 [hep-ex].

[46] I. I. Y. Bigi, M. A. Shifman, N. G. Uraltsev, and A. I. Vainshtein, "The Pole mass of the heavy quark. Perturbation theory and beyond," *Phys. Rev. D* **50** (1994) 2234–2246, arXiv:hep-ph/9402360 [hep-ph].

[47] A. H. Hoang and I. W. Stewart, "Top Mass Measurements from Jets and the Tevatron Top-Quark Mass," *Nucl. Phys. Proc. Suppl.* **185** (2008) 220, arXiv:0808.0222 [hep-ph].

[48] CMS Collaboration, "Determination of the top-quark pole mass and strong coupling constant from the t t-bar production cross section in pp collisions at $\sqrt{s} = 7$ TeV," *Phys. Lett. B* **728** (2014) 496–517, arXiv:1307.1907 [hep-ex]. [Erratum: Phys. Lett.B738,526(2014)].

[49] CMS Collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett. B* **716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[50] D. Kotlinski, E. Bartz, W. Erdmann, K. Gabathuler, R. Horisberger, C. Hormann, H. C. Kastli, B. Meier, and S. Schnetzer, "The control and readout systems of the CMS pixel barrel detector," *Nucl. Instrum. Meth. A* **565** (2006) 73–78.

[51] D. A. Matzner Dominguez, D. Abbaneo, K. Arndt, N. Bacchetta, A. Ball, E. Bartz, W. Bertl, G. M. Bilei, G. Bolla, H. W. K. Cheung, *et al.*, "CMS Technical Design Report for the Pixel Detector Upgrade,".

[52] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at sqrt(s) = 7 TeV at the LHC," *JINST* **8** (2013) P11002, arXiv:1306.6905 [physics.ins-det].

[53] CMS Collaboration, "Commissioning of the Particle-flow Event Reconstruction with the first LHC collisions recorded in the CMS detector,".

[54] CMS Collaboration, "Commissioning of the Particle-Flow reconstruction in Minimum-Bias and Jet Events from pp Collisions at 7 TeV,". https://cds.cern.ch/record/1279341.

[55] CMS Collaboration, "Particle-flow commissioning with muons and electrons from J/Psi and W events at 7 TeV,".

[56] CMS Collaboration, "Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8\,\text{TeV}$," *JINST* **10** (2015) P06005, arXiv:1502.02701 [physics.ins-det].

[57] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at sqrt(s) = 8 TeV," *JINST* **10** no. 08, (2015) P08010, arXiv:1502.02702 [physics.ins-det].

[58] CMS Collaboration, "Performance of CMS muon reconstruction in pp collision events at $\sqrt{s} = 7$ TeV," *JINST* **7** (2012) P10002, arXiv:1206.4071 [physics.ins-det].

[59] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-k_t jet clustering algorithm," *JHEP* **04** (2008) 063, arXiv:0802.1189 [hep-ex].

[60] M. Cacciari, G. P. Salam, and G. Soyez, "FastJet User Manual," *Eur. Phys. J. C* **72** (2012) 1896, arXiv:1111.6097 [hep-ph].

[61] CMS Collaboration, "Missing transverse energy performance of the CMS detector," *JINST* **6** (2011) P09001, arXiv:1106.5048 [physics.ins-det].

[62] CMS Collaboration, "Performance of missing energy reconstruction in 13 TeV pp collision data using the CMS detector,".

[63] CDF Collaboration, "Search for Anomalous Production of Events with Two Photons and Additional Energetic Objects at CDF," *Phys. Rev. D* **82** (2010) 052005, arXiv:0910.5170 [hep-ex].

[64] T. Sjöstrand, S. Mrenna, and P. Skands, "PYTHIA 6.4 physics and manual," *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175 [hep-ph].

[65] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, "MadGraph 5 : Going Beyond," *JHEP* **06** (2011) 128, arXiv:1106.0522 [hep-ph].

[66] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations," *JHEP* **07** (2014) 079, arXiv:1405.0301 [hep-ph].

[67] R. Field, "Early LHC underlying event data – Findings and surprises." 2010.

[68] CMS Collaboration, "Study of the underlying event at forward rapidity in pp collisions at $\sqrt{s} = 0.9$, 2.76, and 7 TeV," *JHEP* **04** (2013) 072, arXiv:1302.2394 [hep-ex].

[69] S. Agostinelli *et al.*, "GEANT4: A Simulation toolkit," *Nucl. Instrum. Meth. A* **506** (2003) 250.

[70] CMS Collaboration, "Search for supersymmetry in pp collisions at $\sqrt{s} = 7$ TeV in events with a single lepton, jets, and missing transverse momentum," *JHEP* **08** (2011) 156, arXiv:1107.1870 [hep-ex].

[71] C. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006. www.GaussianProcess.org/gpml. ISBN 026218253X.

[72] C. Bishop, *Pattern recognition and machine learning*. Springer-Verlag, New York, 1 ed., 2006. ISBN 978-0-387-31073-2.

[73] CMS Collaboration, "Measurement of the top quark mass in the dileptonic ttbar decay channel using the Mbl, MT2, and MAOS Mblv observables,".

[74] A. J. Barr and C. G. Lester, "A Review of the Mass Measurement Techniques proposed for the Large Hadron Collider," *J. Phys. G* **37** (2010) 123001, arXiv:1004.2732 [hep-ph].

[75] C. G. Lester and D. J. Summers, "Measuring masses of semi-invisibly decaying particles pair produced at hadron colliders," *Phys. Lett. B* **463** (1999) 99, arXiv:hep-ph/9906349 [hep-ph].

[76] H.-C. Cheng and Z. Han, "Minimal Kinematic Constraints and m(T2)," *JHEP* **12** (2008) 063, arXiv:0810.5178 [hep-ph].

[77] W. S. Cho, K. Choi, Y. G. Kim, and C. B. Park, "M(T2)-assisted on-shell reconstruction of missing momenta and its application to spin measurement at the LHC," *Phys. Rev. D* **79** (2009) 031701, arXiv:0810.4853 [hep-ph].

[78] CMS Collaboration, "CMS Luminosity Based on Pixel Cluster Counting - Summer 2013 Update,".

[79] CMS Collaboration, "Determination of jet energy calibration and transverse momentum resolution in CMS," *JINST* **6** (2011) P11002, arXiv:1107.4277 [physics.ins-det].

[80] P. Artoisenet, R. Frederix, O. Mattelaer, and R. Rietkerk, "Automatic spin-entangled decays of heavy resonances in Monte Carlo simulations," *JHEP* **03** (2013) 015, arXiv:1212.3460 [hep-ph].

[81] S. Jadach, Z. Was, R. Decker, and J. H. Kuhn, "The tau decay library TAUOLA: Version 2.4," *Comput. Phys. Commun.* **76** (1993) 361.

[82] W.-K. Tung, "New generation of parton distributions with uncertainties from global QCD analysis," *Acta Phys. Polon. B* **33** (2002) 2933, arXiv:hep-ph/0206114 [hep-ph].

[83] S. Alioli, P. Nason, C. Oleari, and E. Re, "NLO single-top production matched with shower in POWHEG: s- and t-channel contributions," *JHEP* **09** (2009) 111, arXiv:0907.4076 [hep-ph]. [Erratum: JHEP02,011(2010)].

[84] E. Re, "Single-top Wt-channel production matched with parton showers using the POWHEG method," *Eur. Phys. J. C* **71** (2011) 1547, arXiv:1009.2450 [hep-ph].

[85] S. Frixione, P. Nason, and C. Oleari, "Matching NLO QCD computations with Parton Shower simulations: the POWHEG method," *JHEP* **11** (2007) 070, arXiv:0709.2092 [hep-ph].

[86] S. Alioli, P. Nason, C. Oleari, and E. Re, "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX," *JHEP* **06** (2010) 043, arXiv:1002.2581 [hep-ph].

[87] N. Kidonakis, "Next-to-next-to-leading soft-gluon corrections for the top quark cross section and transverse momentum distribution," *Phys. Rev. D* **82** (2010) 114030, arXiv:1009.4935 [hep-ph].

[88] N. Kidonakis, "Differential and total cross sections for top pair and single top production," in *Proceedings, 20th International Workshop on Deep-Inelastic Scattering and Related Subjects (DIS 2012)*, p. 831. 2012. arXiv:1205.3453 [hep-ph]. https://inspirehep.net/record/1114754/files/arXiv:1205.3453.pdf.

[89] K. Melnikov and F. Petriello, "Electroweak gauge boson production at hadron colliders through O(alpha(s)**2)," *Phys. Rev. D* **74** (2006) 114017, arXiv:hep-ph/0609070 [hep-ph].

[90] J. M. Campbell and R. K. Ellis, "MCFM for the Tevatron and the LHC," *Nucl. Phys. Proc. Suppl.* **205-206** (2010) 10, arXiv:1007.3492 [hep-ph].

[91] M. Czakon, P. Fiedler, and A. Mitov, "Total Top-Quark Pair-Production Cross Section at Hadron Colliders Through $O(\frac{4}{S})$," *Phys. Rev. Lett.* **110** (2013) 252004, arXiv:1303.6254 [hep-ph].

[92] M. Burns, K. Kong, K. T. Matchev, and M. Park, "Using Subsystem MT2 for Complete Mass Determinations in Decay Chains with Missing Energy at Hadron Colliders," *JHEP* **03** (2009) 143, arXiv:0810.5576 [hep-ph].

[93] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. Monographs on Statistics and Applied Probability. Chapman & Hall, 1993.

[94] CMS and ATLAS Collaborations, "Jet energy scale uncertainty correlations between ATLAS and CMS at 8 TeV," Tech. Rep. CMS-PAS-JME-15-001, ATL-PHYS-PUB-2015-049, CERN, Geneva, 2015. http://cds.cern.ch/record/2104039.

[95] M. Bahr *et al.*, "Herwig++ Physics and Manual," *Eur. Phys. J. C* **58** (2008) 639, arXiv:0803.0883 [hep-ph].

[96] CMS Collaboration, "Identification of b-quark jets with the CMS experiment," *JINST* **8** (2013) P04013, arXiv:1211.4462 [hep-ex].

[97] CMS Collaboration, "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8\,\text{TeV}$," *Eur. Phys. J. C* **75** (2015) 542, arXiv:1505.04480 [hep-ex].

[98] P. Z. Skands, "Tuning Monte Carlo Generators: The Perugia Tunes," *Phys. Rev. D* **82** (2010) 074018, arXiv:1005.3457 [hep-ph].

[99] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P. M. Nadolsky, J. Pumplin, and C. P. Yuan, "New parton distributions for collider physics," *Phys. Rev. D* **82** (2010) 074024, arXiv:1007.2241 [hep-ph].

[100] ALEPH Collaboration, "Study of the fragmentation of b quarks into B mesons at the Z peak," *Phys. Lett. B* **512** (2001) 30, arXiv:hep-ex/0106051 [hep-ex].

[101] DELPHI Collaboration, "A study of the b-quark fragmentation function with the DELPHI detector at LEP I and an averaged distribution obtained at the Z Pole," *Eur. Phys. J. C* **71** (2011) 1557, arXiv:1102.4748 [hep-ex].

[102] H. Crámer, *Mathematical Methods of Statistics*. Princeton University Press, 1999.

[103] A. DasGupta, *Asymptotic Theory of Statistics and Probability*. Springer, 2008.

[104] T. Teubner, "Physics of the top quark at future lepton colliders," *Journal of Physics: Conference Series* **452** no. 1, (2013) 012043. http://stacks.iop.org/1742-6596/452/i=1/a=012043.

[105] M. Vos *et al.*, "Top physics at high-energy lepton colliders," arXiv:1604.08122 [hep-ex].

[106] D. d'Enterria, "Physics at the FCC-ee," in *17th Lomonosov Conference on Elementary Particle Physics Moscow, Russia, August 20-26, 2015*. 2016. arXiv:1602.05043 [hep-ex]. https://inspirehep.net/record/1421932/files/arXiv:1602.05043.pdf.

[107] R. Brun and F. Rademakers, "ROOT - An Object Oriented Data Analysis Framework," in *Proceedings AIHENP'96 Workshop, Lausanne, Sep. 1996, Nucl. Inst. & Meth. in Phys. Res. A 389*, p. 81. 1997. http://root.cern.ch/.

[108] F. James and M. Roos, "Minuit: A System for Function Minimization and Analysis of the Parameter Errors and Correlations," *Comput. Phys. Commun.* **10** (1975) 343–367.